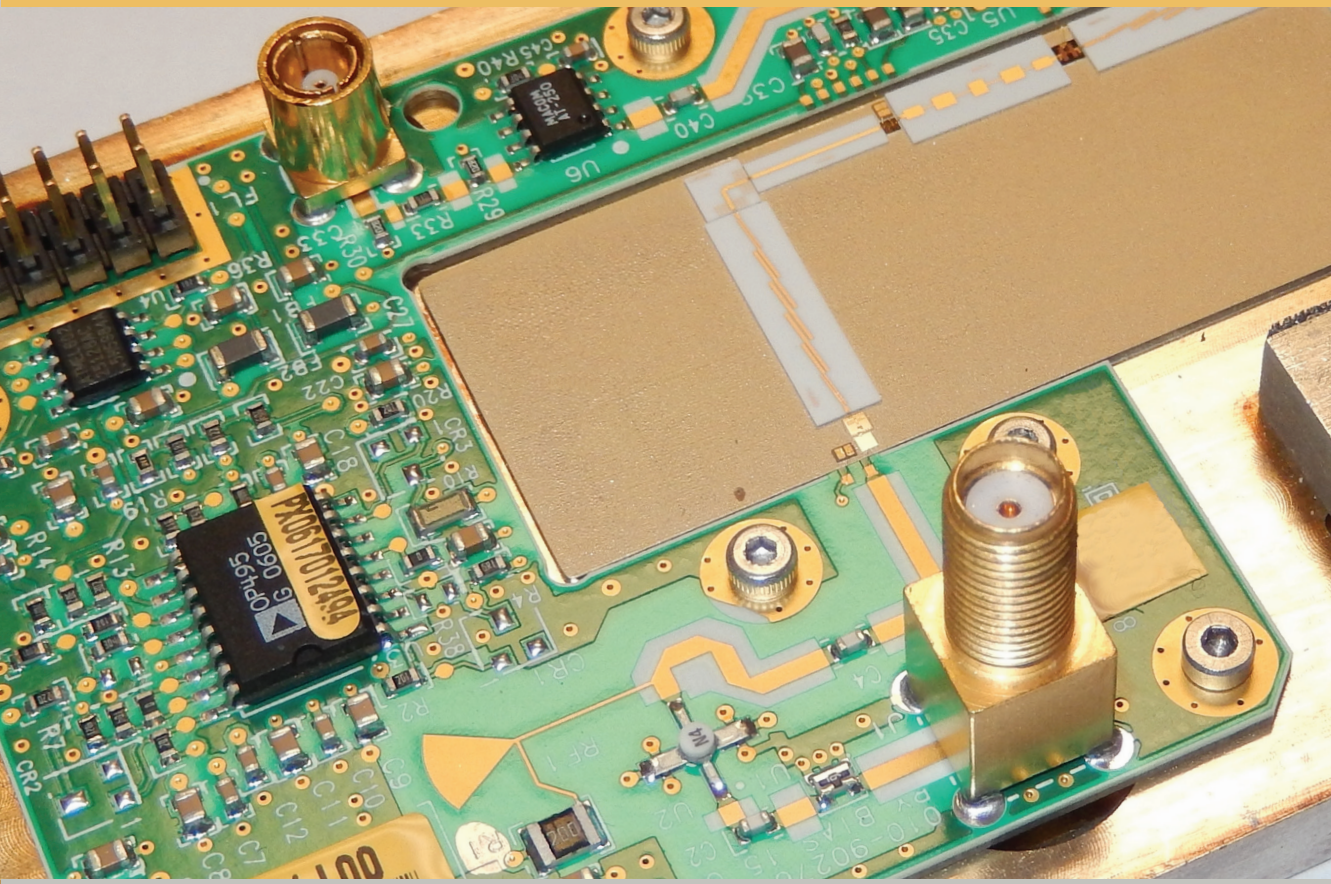
Volume 4

MICROWAVE AND RF DESIGN MODULES



Michael Steer

Third Edition

Microwave and RF Design
Modules

Volume 4

Third Edition

Michael Steer

Microwave and RF Design
Modules

Volume 4

Third Edition

Michael Steer

Copyright © 2019 by M.B. Steer

Citation: Steer, Michael. *Microwave and RF Design: Modules*. Volume 4. (Third Edition), NC State University, 2019. doi: https//doi.org/10.5149/9781469656977 Steer

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (CC BY-NC 4.0). To view a copy of the license, visit http://creativecommons.org/licenses.

ISBN 978-1-4696-5696-0 (paperback)
ISBN 978-1-4696-5697-7 (open access ebook)

Published by NC State University

NC STATE UNIVERSITY

Distributed by the University of North Carolina Press
www.uncpress.org

Printing: 1

To my Son- and Daughter-in-Laws

Claudio and Tracy

Preface

The book series *Microwave and RF Design* is a comprehensive treatment of radio frequency (RF) and microwave design with a modern "systems-first" approach. A strong emphasis on design permeates the series with extensive case studies and design examples. Design is oriented towards cellular communications and microstrip design so that lessons learned can be applied to real-world design tasks. The books in the Microwave and RF Design series are:

- Microwave and RF Design: Radio Systems, Volume 1
- Microwave and RF Design: Transmission Lines, Volume 2
- Microwave and RF Design: Networks, Volume 3
- Microwave and RF Design: Modules, Volume 4
- Microwave and RF Design: Amplifiers and Oscillators, Volume 5

The length and format of each is suitable for automatic printing and binding.

Rationale

The central philosophy behind this series's popular approach is that the student or practicing engineer will develop a full appreciation for RF and microwave engineering and gain the practical skills to perform system-level design decisions. Now more than ever companies need engineers with an ingrained appreciation of systems and armed with the skills to make system decisions. One of the greatest challenges facing RF and microwave engineering is the increasing level of abstraction needed to create innovative microwave and RF systems. This book series is organized in such a way that the reader comes to understand the impact that system-level decisions have on component and subsystem design. At the same time, the capabilities of technologies, components, and subsystems impact system design. The book series is meticulously crafted to intertwine these themes.

Audience

The book series was originally developed for three courses at North Carolina State University. One is a final-year undergraduate class, another an introductory graduate class, and the third an advanced graduate class. Books in the series are used as supplementary texts in two other classes. There are extensive case studies, examples, and end of chapter problems ranging from straight-forward to in-depth problems requiring hours to solve. A companion book, *Fundamentals of Microwave and RF Design*, is more suitable for an undergraduate class yet there is a direct linkage between the material in this book and the series which can then be used as a career-long reference text. I believe it is completely understandable for senior-level students where a microwave/RF engineering course is offered. The book series is a comprehensive RF and microwave text and reference, with detailed index, appendices, and cross-references throughout. Practicing engineers will find the book series a valuable systems primer, a refresher as needed, and a

reference tool in the field. Additionally, it can serve as a valuable, accessible resource for those outside RF circuit engineering who need to understand how they can work with RF hardware engineers.

Organization

This book is a volume in a five volume series on RF and microwave design. The first volume in the series, *Microwave and RF Design: Radio Systems*, addresses radio systems mainly following the evolution of cellular radio. A central aspect of microwave engineering is distributed effects considered in the second volume of this book series, *Microwave and RF Design: Transmission Lines*. Here transmission lines are treated as supporting forward- and backward-traveling voltage and current waves and these are related to electromagnetic effects. The third volume, *Microwave and RF Design: Networks*, covers microwave network theory which is the theory that describes power flow and can be used with transmission line effects. Topics covered in *Microwave and RF Design: Modules*, focus on designing microwave circuits and systems using modules introducing a large number of different modules. Modules is just another term for a network but the implication is that is is packaged and often available off-the-shelf. Other topics that are important in system design using modules are considered including noise, distortion, and dynamic range. Most microwave and RF designers construct systems using modules developed by other engineers who specialize in developing the modules. Examples are filter and amplifier modules which once designed can be used in many different systems. Much of microwave design is about maximizing dynamic range, minimizing noise, and minimizing DC power consumption. The fifth volume in this series, *Microwave and RF Design: Amplifiers and Oscillators*, considers amplifier and oscillator design and develops the skills required to develop modules.

Volume 1: Microwave and RF Design: Radio Systems

The first book of the series covers RF systems. It describes system concepts and provides comprehensive knowledge of RF and microwave systems. The emphasis is on understanding how systems are crafted from many different technologies and concepts. The reader gains valuable insight into how different technologies can be traded off in meeting system requirements. I do not believe this systems presentation is available anywhere else in such a compact form.

Volume 2: Microwave and RF Design: Transmission Lines

This book begins with a chapter on transmission line theory and introduces the concepts of forward- and backward-traveling waves. Many examples are included of advanced techniques for analyzing and designing transmission line networks. This is followed by a chapter on planar transmission lines with microstrip lines primarily used in design examples. Design examples illustrate some of the less quantifiable design decisions that must be made. The next chapter describes frequency-dependent transmission line effects and describes the design choices that must be taken to avoid multimoding. The final chapter in this volume addresses coupled-lines. It is shown how to design coupled-line networks that exploit this distributed effect to realize novel circuit functionality and how to design networks that minimize negative effects. The modern treatment of transmission lines in this volume emphasizes planar circuit design and the practical aspects of designing

around unwanted effects. Detailed design of a directional coupler is used to illustrate the use of coupled lines. Network equivalents of coupled lines are introduced as fundamental building blocks that are used later in the synthesis of coupled-line filters. The text, examples, and problems introduce the often hidden design requirements of designing to mitigate parasitic effects and unwanted modes of operation.

Volume 3: Microwave and RF Design: Networks

Volume 3 focuses on microwave networks with descriptions based on S parameters and $ABCD$ matrices, and the representation of reflection and transmission information on polar plots called Smith charts. Microwave measurement and calibration technology are examined. A sampling of the wide variety of microwave elements based on transmission lines is presented. It is shown how many of these have lumped-element equivalents and how lumped elements and transmission lines can be combined as a compromise between the high performance of transmission line structures and the compactness of lumped elements. This volume concludes with an in-depth treatment of matching for maximum power transfer. Both lumped-element and distributed-element matching are presented.

Volume 4: Microwave and RF Design: Modules

Volume 4 focuses on the design of systems based on microwave modules. The book considers the wide variety of RF modules including amplifiers, local oscillators, switches, circulators, isolators, phase detectors, frequency multipliers and dividers, phase-locked loops, and direct digital synthesizers. The use of modules has become increasingly important in RF and microwave engineering. A wide variety of passive and active modules are available and high-performance systems can be realized cost effectively and with stellar performance by using off-the-shelf modules interconnected using planar transmission lines. Module vendors are encouraged by the market to develop competitive modules that can be used in a wide variety of applications. The great majority of RF and microwave engineers either develop modules or use modules to realize RF systems. Systems must also be concerned with noise and distortion, including distortion that originates in supposedly linear elements. Something as simple as a termination can produce distortion called passive intermodulation distortion. Design techniques are presented for designing cascaded systems while managing noise and distortion. Filters are also modules and general filter theory is covered and the design of parallel coupled line filters is presented in detail. Filter design is presented as a mixture of art and science. This mix, and the thought processes involved, are emphasized through the design of a filter integrated throughout this chapter.

Volume 5: Microwave and RF Design: Amplifiers and Oscillators

The fifth volume presents the design of amplifiers and oscillators in a way that enables state-of-the-art designs to be developed. Detailed strategies for amplifiers and voltage-controlled oscillators are presented. Design of competitive microwave amplifiers and oscillators are particularly challenging as many trade-offs are required in design, and the design decisions cannot be reduced to a formulaic flow. Very detailed case studies are presented and while some may seem quite complicated, they parallel the level of sophistication required to develop competitive designs.

Case Studies

A key feature of this book series is the use of real world case studies of leading edge designs. Some of the case studies are designs done in my research group to demonstrate design techniques resulting in leading performance. The case studies and the persons responsible for helping to develop them are as follows.

1. Software defined radio transmitter.
2. High dynamic range down converter design. This case study was developed with Alan Victor.
3. Design of a third-order Chebyshev combline filter. This case study was developed with Wael Fathelbab.
4. Design of a bandstop filter. This case study was developed with Wael Fathelbab.
5. Tunable Resonator with a varactor diode stack. This case study was developed with Alan Victor.
6. Analysis of a 15 GHz Receiver. This case study was developed with Alan Victor.
7. Transceiver Architecture. This case study was developed with Alan Victor.
8. Narrowband linear amplifier design. This case study was developed with Dane Collins and National Instruments Corporation.
9. Wideband Amplifier Design. This case study was developed with Dane Collins and National Instruments Corporation.
10. Distributed biasing of differential amplifiers. This case study was developed with Wael Fathelbab.
11. Analysis of a distributed amplifier. This case study was developed with Ratan Bhatia, Jason Gerber, Tony Kwan, and Rowan Gilmore.
12. Design of a WiMAX power amplifier. This case study was developed with Dane Collins and National Instruments Corporation.
13. Reflection oscillator. This case study was developed with Dane Collins and National Instruments Corporation.
14. Design of a C-Band VCO. This case study was developed with Alan Victor.
15. Oscillator phase noise analysis. This case study was developed with Dane Collins and National Instruments Corporation.

Many of these case studies are available as captioned YouTube videos and qualified instructors can request higher resolution videos from the author.

Course Structures

Based on the adoption of the first and second editions at universities, several different university courses have been developed using various parts of what was originally one very large book. The book supports teaching two or three classes with courses varying by the selection of volumes and chapters. A standard microwave class following the format of earlier microwave texts can be taught using the second and third volumes. Such a course will benefit from the strong practical design flavor and modern treatment of measurement technology, Smith charts, and matching networks. Transmission line propagation and design is presented in the context of microstrip technology providing an immediately useful skill. The subtleties of multimoding are also presented in the context of microstrip lines. In such

a class the first volume on microwave systems can be assigned for self-learning.

Another approach is to teach a course that focuses on transmission line effects including parallel coupled-line filters and module design. Such a class would focus on Volumes 2, 3 and 4. A filter design course would focus on using Volume 4 on module design. A course on amplifier and oscillator design would use Volume 5. This course is supported by a large number of case studies that present design concepts that would otherwise be difficult to put into the flow of the textbook.

Another option suited to an undergraduate or introductory graduate class is to teach a class that enables engineers to develop RF and microwave systems. This class uses portions of Volumes 2, 3 and 4. This class then omits detailed filter, amplifier, and oscillator design.

The fundamental philosophy behind the book series is that the broader impact of the material should be presented first. Systems should be discussed up front and not left as an afterthought for the final chapter of a textbook, the last lecture of the semester, or the last course of a curriculum.

The book series is written so that all electrical engineers can gain an appreciation of RF and microwave hardware engineering. The body of the text can be covered without strong reliance on this electromagnetic theory, but it is there for those who desire it for teaching or reader review. The book is rich with detailed information and also serves as a technical reference.

The Systems Engineer

Systems are developed beginning with fuzzy requirements for components and subsystems. Just as system requirements provide impetus to develop new base technologies, the development of new technologies provides new capabilities that drive innovation and new systems. The new capabilities may arise from developments made in support of other systems. Sometimes serendipity leads to the new capabilities. Creating innovative microwave and RF systems that address market needs or provide for new opportunities is the most exciting challenge in RF design. The engineers who can conceptualize and architect new RF systems are in great demand. This book began as an effort to train RF systems engineers and as an RF systems resource for practicing engineers. Many RF systems engineers began their careers when systems were simple. Today, appreciating a system requires higher levels of abstraction than in the past, but it also requires detailed knowledge or the ability to access detailed knowledge and expertise. So what makes a systems engineer? There is not a simple answer, but many partial answers. We know that system engineers have great technical confidence and broad appreciation for technologies. They are both broad in their knowledge of a large swath of technologies and also deep in knowledge of a few areas, sometimes called the “T” model. One book or course will not make a systems engineer. It is clear that there must be a diverse set of experiences. This book series fulfills the role of fostering both high-level abstraction of RF engineering and also detailed design skills to realize effective RF and microwave modules. My hope is that this book will provide the necessary background for the next generation of RF systems engineers by stressing system principles immediately, followed by core RF technologies. Core technologies are thereby covered within the context of the systems in which they are used.

Supplementary Materials

Supplementary materials available to qualified instructors adopting the book include PowerPoint slides and solutions to the end-of-chapter problems. Requests should be directed to the author. Access to downloads of the books, additional material and YouTube videos of many case studies are available at `https://www.lib.ncsu.edu/do/open-education`

Acknowledgments

Writing this book has been a large task and I am indebted to the many people who helped along the way. First I want to thank the more than 1200 electrical engineering graduate students who used drafts and the first two editions at NC State. I thank the many instructors and students who have provided feedback. I particularly thank Dr. Wael Fathelbab, a filter expert, who co-wrote an early version of the filter chapter. Professor Andreas Cangellaris helped in developing the early structure of the book. Many people have reviewed the book and provided suggestions. I thank input on the structure of the manuscript: Professors Mark Wharton and Nuno Carvalho of Universidade de Aveiro, Professors Ed Delp and Saul Gelfand of Purdue University, Professor Lynn Carpenter of Pennsylvania State University, Professor Grant Ellis of the Universiti Teknologi Petronas, Professor Islam Eshrah of Cairo University, Professor Mohammad Essaaidi and Dr. Otman Aghzout of Abdelmalek Essaadi Univeristy, Professor Jianguo Ma of Guangdong University of Technology, Dr. Jayesh Nath of Apple, Mr. Sony Rowland of the U.S. Navy, and Dr. Jonathan Wilkerson of Lawrence Livermore National Laboratories, Dr. Josh Wetherington of Vadum, Dr. Glen Garner of Vadum, and Mr. Justin Lowry who graduated from North Carolina State University.

Many people helped in producing this book. In the first edition I was assisted by Ms. Claire Sideri, Ms. Susan Manning, and Mr. Robert Lawless who assisted in layout and production. The publisher, task master, and chief coordinator, Mr. Dudley Kay, provided focus and tremendous assistance in developing the first and second editions of the book, collecting feedback from many instructors and reviewers. I thank the Institution of Engineering and Technology, who acquired the original publisher, for returning the copyright to me. This open access book was facilitated by John McLeod and Samuel Dalzell of the University of North Carolina Press, and by Micah Vandergrift and William Cross of NC State University Libraries. The open access ebooks are host by NC State University Libraries.

The book was produced using LaTeX and open access fonts, line art was drawn using xfig and inkscape, and images were edited in gimp. So thanks to the many volunteers who developed these packages.

My family, Mary, Cormac, Fiona, and Killian, gracefully put up with my absence for innumerable nights and weekends, many more than I could have ever imagined. I truly thank them. I also thank my academic sponsor, Dr. Ross Lampe, Jr., whose support of the university and its mission enabled me to pursue high risk and high reward endeavors including this book.

Michael Steer
North Carolina State University
Raleigh, North Carolina
mbs@ncsu.edu

List of Trademarks

3GPP® is a registered trademark of the European Telecommunications Standards Institute.
802® is a registered trademark of the Institute of Electrical & Electronics Engineers .
APC-7® is a registered trademark of Amphenol Corporation.
AT&T® is a registered trademark of AT&T Intellectual Property II, L.P.
AWR® is a registered trademark of National Instruments Corporation.
AWRDE® is a trademark of National Instruments Corporation.
Bluetooth® is a registered trademark of the Bluetooth Special Interest Group.
GSM® is a registered trademark of the GSM MOU Association.
Mathcad® is a registered trademark of Parametric Technology Corporation.
MATLAB® is a registered trademark of The MathWorks, Inc.
NEC® is a registered trademark of NEC Corporation.
OFDMA® is a registered trademark of Runcom Technologies Ltd.
Qualcomm® is a registered trademark of Qualcomm Inc.
Teflon® is a registered trademark of E. I. du Pont de Nemours.
RFMD® is a registered trademark of RF Micro Devices, Inc.
SONNET® is a trademark of Sonnet Corporation.
Smith is a registered trademark of the Institute of Electrical and Electronics Engineers.
Touchstone® is a registered trademark of Agilent Corporation.
WiFi® is a registered trademark of the Wi-Fi Alliance.
WiMAX® is a registered trademark of the WiMAX Forum.

All other trademarks are the properties of their respective owners.

Contents

Preface v

1 Introduction to RF and Microwave Modules 1
1.1 Introduction to Microwave Modules 1
1.2 A 15 GHz Receiver Subsystem 2
1.3 Book Outline 2
1.4 References 4
1.A RF and Microwave Circuit Schematic Symbols 5
1.A.1 Element and Circuit Symbols 5
1.A.2 Sources 8
1.A.3 Diodes 8
1.A.4 Bipolar Junction Transistor 8
1.A.5 Junction Field Effect Transistor 9
1.A.6 Metal-Oxide-Semiconductor Field Effect Transistor . . 9

2 Filters 11
2.1 Introduction 11
2.1.1 Filter Prototypes 12
2.1.2 Image Parameter Versus Insertion loss Methods 13
2.2 Singly and Doubly Terminated Networks 13
2.2.1 Doubly Terminated Networks 14
2.2.2 Lowpass Filter Response 15
2.3 The Lowpass Filter Prototype 17
2.4 The Maximally Flat (Butterworth) Lowpass Approximation . 18
2.4.1 Butterworth Filter Design 18
2.4.2 Construction of the Transfer Function 20
2.4.3 nth-Order Reflection Approximation 21
2.4.4 Bandwidth Consideration 21
2.5 The Chebyshev Lowpass Approximation 22
2.5.1 Chebyshev Filter Design 22
2.5.2 Chebyshev Approximation and Recursion 23
2.5.3 Bandwidth Consideration 24
2.6 Element Extraction 24
2.6.1 Ladder Synthesis 24
2.6.2 Summary 28
2.7 Butterworth and Chebyshev Filters 29
2.7.1 Butterworth Filter 30
2.7.2 Chebyshev Filter 31
2.7.3 Summary 32
2.8 Impedance and Admittance Inverters 33
2.8.1 Properties of an Impedance Inverter 34
2.8.2 Replacement of a Series Inductor by a Shunt Capacitor 34
2.8.3 Replacement of a Series Capacitor by a Shunt Inductor 36
2.8.4 Ladder Prototype with Impedance Inverters 36

2.8.5 Lumped-Element Realization of an Inverter 37
2.8.6 Narrowband Realization of an Inverter Using Transmission Line Stubs 39
2.9 Filter Transformations 40
2.9.1 Impedance Transformation 40
2.9.2 Frequency Transformation: Lowpass 41
2.9.3 Lowpass to Highpass Transformation 42
2.9.4 Lowpass to Bandpass Transformation 42
2.9.5 Lowpass to Bandstop Transformation 45
2.9.6 Transformed Ladder Prototypes 46
2.10 Cascaded Line Realization of Filters 46
2.11 Butterworth and Chebyshev Bandpass Filters 48
2.12 Richards's Transformation 50
2.12.1 Richards's Transformation and Transmission Lines . . 51
2.12.2 Richards's Transformation and Stubs 52
2.12.3 Richards's Transformation Applied to a Lowpass Filter 52
2.12.4 Richards's Transformation Applied to a Highpass Filter 53
2.13 Kuroda's and Norton's Network Identities 54
2.13.1 Kuroda's Identities 54
2.13.2 Norton's Identities 56
2.14 Inter-resonator Coupled Bandpass Filters 57
2.15 Bandpass Filter Topologies 61
2.16 Case Study: Design of a Bandstop Filter 62
2.17 Active Filters 64
2.17.1 Radio Frequency Active Filters 66
2.17.2 Biquadratic Filters 67
2.17.3 Distributed Active Filters 68
2.18 Transient Response of a Bandpass Filter 69
2.19 Summary 71
2.20 References 72
2.21 Exercises 74
2.21.1 Exercises by Section 77
2.21.2 Answers to Selected Exercises 77

3 Parallel Coupled-Line Filters 79
3.1 Introduction 79
3.2 Parallel Coupled Lines 80
3.2.1 Coupled-Line Configurations 83
3.2.2 Coupled-Line Circuit Models 87
3.3 Inverter Network Scaling 91
3.4 Case Study: Third-Order Chebyshev Combline Filter Design . 93
3.4.1 Realization of the Input/Output Inverters 101
3.4.2 Implementation 102
3.4.3 Alternative Combline Filter Layouts 112
3.5 Parallel Coupled-Line Filters in an Inhomogeneous Medium . 114
3.6 Summary 114
3.7 References 115
3.8 Exercises 116
3.8.1 Exercises by Section 120
3.8.2 Answers to Selected Exercises 120

4 Noise, Distortion, and Dynamic Range . . . 123
4.1 Introduction . . . 123
4.2 Noise . . . 123
4.2.1 Observations of Noise Spectra . . . 124
4.2.2 Characterization of Thermal Noise . . . 126
4.2.3 Environmental Noise . . . 128
4.2.4 Thermal Noise and Capacitors . . . 129
4.2.5 Physical Source of Shot Noise . . . 129
4.2.6 Physical Source of Flicker Noise . . . 130
4.3 Noise Characterization . . . 130
4.3.1 Noise in a Cascaded System . . . 133
4.3.2 Measurement of Noise Figure . . . 135
4.3.3 Measurement of Noise Temperature . . . 137
4.3.4 Measuring the Noise Figure of Low Noise Devices . . . 137
4.3.5 Radiometer System . . . 138
4.3.6 Noise Figure of a Two-Port Amplifier . . . 139
4.4 Oscillator Noise . . . 141
4.4.1 Observations of Noise Spectra of Oscillators . . . 141
4.5 Nonlinear Distortion . . . 144
4.5.1 Amplitude and Phase Distortion . . . 144
4.5.2 Gain Compression . . . 145
4.5.3 Intermodulation Distortion . . . 146
4.5.4 Third-Order Distortion . . . 147
4.5.5 Spectral Regrowth . . . 149
4.5.6 Second-Order Distortion . . . 150
4.5.7 Summary . . . 150
4.6 Dynamic Range . . . 151
4.7 Passive Intermodulation Distortion . . . 154
4.7.1 Sources of PIM . . . 154
4.7.2 Summary . . . 156
4.8 Breakdown . . . 156
4.8.1 Multipactor Effect . . . 157
4.8.2 Corona Effect . . . 157
4.8.3 Summary . . . 158
4.9 Summary . . . 158
4.10 References . . . 159
4.11 Exercises . . . 161
4.11.1 Exercises by Section . . . 163
4.11.2 Answers to Selected Exercises . . . 163

5 Passive Modules . . . 165
5.1 Introduction to Passive Modules . . . 165
5.2 Diodes . . . 167
5.3 Case Study: Tunable Resonator with Varactor Diode Stack . . . 170
5.4 Switch . . . 172
5.5 Ferrite Components: Circulators and Isolators . . . 177
5.5.1 Gyromagnetic Effect . . . 177
5.5.2 Circulator . . . 179
5.5.3 Circulator Isolation . . . 179
5.5.4 Isolator . . . 180
5.5.5 YIG-Tuned Bandpass Filter . . . 181

5.6 Summary . . . 181
5.7 References . . . 182
5.8 Exercises . . . 182
5.8.1 Exercises By Section . . . 184
5.8.2 Answers to Selected Exercises . . . 184

6 Mixer and Source Modules . . . 185
6.1 Introduction . . . 185
6.2 Mixer . . . 186
6.2.1 Mixer Analysis . . . 187
6.2.2 Mixer Performance Parameters . . . 189
6.2.3 Mixer Waveforms . . . 191
6.2.4 Switching Mixer . . . 194
6.2.5 Subharmonic Mixer . . . 196
6.3 Single-Ended, Balanced, and Double Balanced Mixers . . . 197
6.3.1 RFIC Mixers . . . 197
6.3.2 Gilbert Cell . . . 198
6.3.3 Integrated Gilbert Cell-Based Mixer . . . 198
6.3.4 Ring-Based Mixers . . . 200
6.3.5 Summary . . . 201
6.4 Local Oscillator . . . 202
6.4.1 Phase Noise in Local Oscillators . . . 202
6.5 Voltage-Controlled Oscillator (VCO) . . . 204
6.6 Phase Detector . . . 206
6.7 Frequency Multiplier . . . 207
6.8 Frequency Divider . . . 208
6.9 Phase-Locked Loop (PLL) . . . 209
6.9.1 Operation . . . 209
6.9.2 First-Order PLL . . . 210
6.9.3 Applications . . . 211
6.10 Direct Digital Synthesizer . . . 213
6.11 Diode and Vacuum Sources . . . 214
6.11.1 Two-Terminal Semiconductor Sources . . . 214
6.11.2 Vacuum Devices . . . 216
6.12 Summary . . . 217
6.13 References . . . 217
6.14 Exercises . . . 220
6.14.1 Exercises By Section . . . 221
6.14.2 Answers to Selected Exercises . . . 221

7 Cascade of Modules . . . 223
7.1 Introduction . . . 223
7.2 Nonlinear Distortion in a Cascaded System . . . 225
7.2.1 Gain Compression in a Cascaded System . . . 225
7.2.2 Intermodulation Distortion in a Cascaded System . . . 228
7.3 Cascaded Module Design Using the Budget Method . . . 231
7.4 Cascaded Module Design Using the Contribution Method . . . 232
7.4.1 Noise Contribution . . . 233
7.4.2 Intermodulation Contribution . . . 234
7.4.3 Design Methodology for Maximizing Dynamic Range 235
7.4.4 Summary . . . 236

7.5 Case Study: High Dynamic Range Down-Converter Design . 236
7.5.1 Architecture 237
7.5.2 Design 237
7.6 Case Study: Analysis of a 15 GHz Receiver 238
7.7 Case Study: Frequency Planning of a Transceiver 241
7.7.1 Transceiver Architecture 242
7.7.2 Frequency Planning 244
7.7.3 Summary 247
7.8 Summary 247
7.9 References 248
7.10 Exercises 248
7.10.1 Exercises By Section 250
7.10.2 Answers to Selected Exercises 250

Index 251

CHAPTER 1

Introduction to RF and Microwave Modules

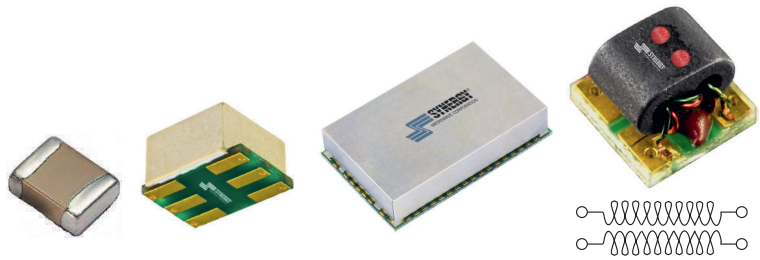
1.1 Introduction to Microwave Modules 1
1.2 A 15 GHz Receiver Subsystem 2
1.3 Book Outline 2
1.4 References 4
1.A Appendix: RF and Microwave Circuit Schematic Symbols 5

1.1 Introduction to Microwave Modules

Most microwave design develops microwave systems using modules such amplifiers, integrated circuits (ICs), filters, frequency multipliers, and passive components. Economics necessitate that since modules are expensive to design, e.g RF integrated circuits, that they be developed for multiple applications. In system design modules are chosen for their dynamic range, noise performance, DC power consumption, and cost. Foremost the system designer must have knowledge of available modules and be prepared to design a module itself if this results in competitive performance and manages cost. Modules are interconnected by transmission lines, bias settings and matching networks must be designed, and a system frequency plan must be developed that trades-off cost and performance while avoiding interference. All these topics are addressed on this book which develops the knowledge of modules, and the skills required to design microwave systems.

Most RF systems are composed of a cascade of modules each of which is separately designed and characterized. They often have $50\ \Omega$ input and output impedances so that modules can be freely interconnected. Just as often matching networks must be design by the user as otherwise operating frequencies are pre-determined by the module vendor as a matching network usually has a narrower bandwidth than that of the functional core of the module. In high volume applications, several modules could be monolithically integrated, but even then, design is based on the concept of modules. Many modules are available "off the shelf" and high-performance RF systems can be constructed using them.

Examples of commercially available modules are shown in Figure 1-1. They can range in complexity from the surface mount resistor of Figure 1-1(a) and the transformer of Figure 1-1(d), up to the mixer and synthesizer modules shown in Figures 1-1(b and c). Modules can have very good performance as it is cost effective to put considerable design effort into a module that can be used in many applications and thus design costs shared.



(a) Resistor (b) Mixer (c) Synthesizer (d) Transformer

Figure 1-1: Modules in surface-mount packages. Copyright Synergy Microwave Corporation, used with permission [1].

1.2 A 15 GHz Receiver Subsystem

An example of a microwave subsystem designed using modules is the 15 GHz receiver subsystem shown in Figure 1-2 with details of the frequency conversion section shown in Figure 1-3. This receiver subsystem is itself a module used in building a complete transceiver system for a point-to-point microwave link connecting two basestations. The modules of the receiver include amplifiers, frequency multipliers, a mixers, an isolator, and a waveguide adaptor. These modules are available as off-the-shelf components from companies that specialize in developing them, and selling them to a large user base. This receiver is the subject of a case study in Section 7.6 where more details of the modules, frequency planning, and signal flow are given.

1.3 Book Outline

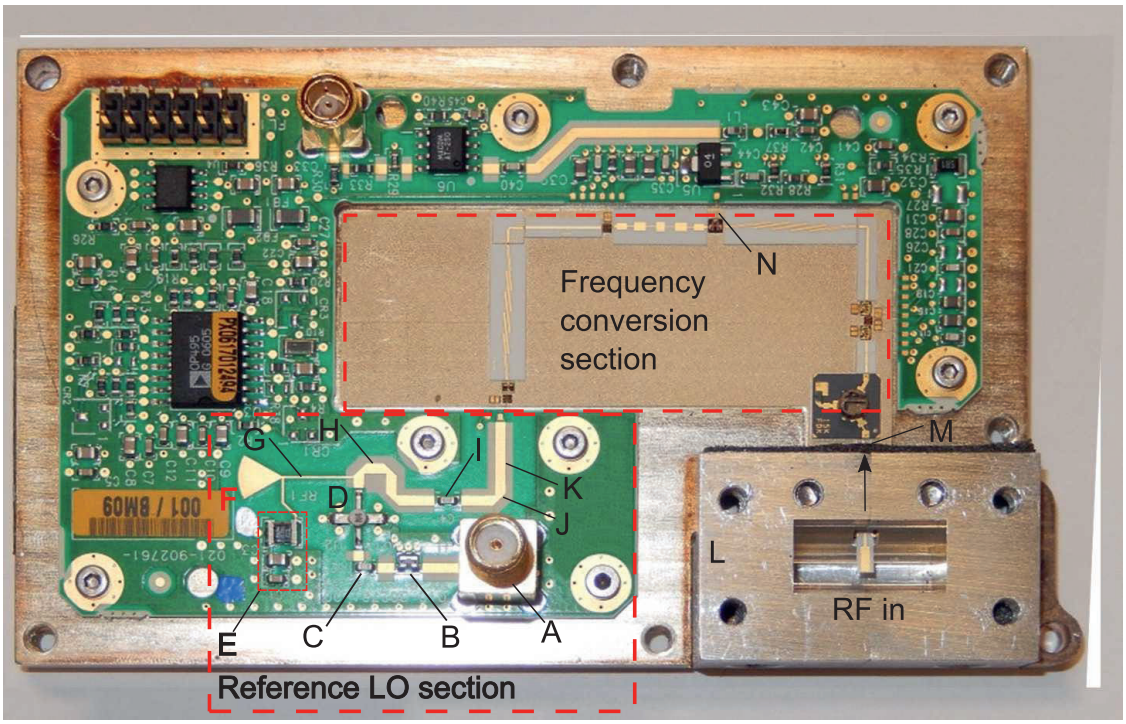
This book describes modules and tools required to design microwave systems using modules. A large number of modules are presented with the aim of building a 'module vocabulary.' An engineer must know what modules are available before a microwave system can be designed using modules.

The first set of modules examined are filters in Chapter 2. These have the important task of only letting wanted signals through a circuit while blocking noise and unwanted signals. Then parallel coupled line (PCL) filters are considered in Chapter 3. A PCL filter is the most common type of microwave filter and is based on the inherent filtering properties of a pair of coupled transmission lines.

Chapter 4 introduces metrics for noise, distortion, and dynamic range and introduces the tools that are required to design systems of cascaded modules that minimize noise and distortion, and maximize dynamic range.

Chapter 5 develops knowledge of passive modules, particular ones that use diodes and switches. The final chapter, Chapter 6, considers mixer and source modules.

This book is the fourth volume in a series on microwave and RF design. The first volume in the series addresses radio systems [2] mainly following the evolution of cellular radio. A central aspect of microwave engineering is distributed effects considered in the second volume of the series [3]. Here the transmission lines are treated as supporting forward- and backward-traveling voltage and current waves and these are related to electromagnetic effects. The third volume [4] covers microwave network theory which is the theory that describes power flow in microwave circuits and can be used to describe transmission line effects. Topics covered in this volume include scattering parameters, Smith charts, and matching networks that



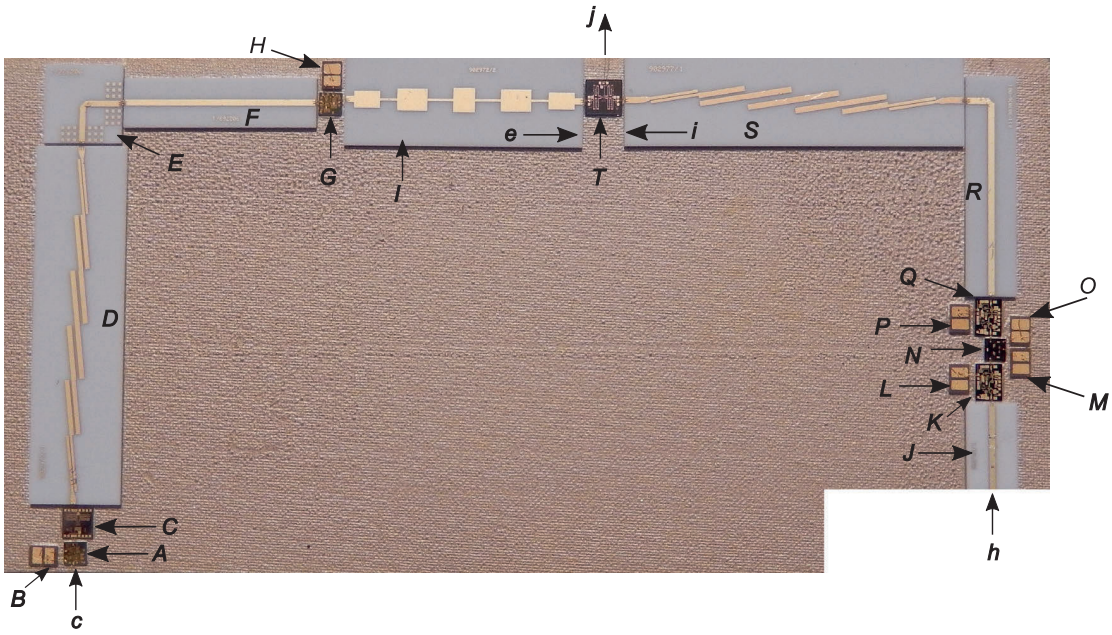
A	SMA connector, Reference LO in	**H**	50 Ω transmission line
B	Attenuator	**I**	DC blocking capacitor
C	DC blocking capacitor	**J**	Mitered bend
D	Reference LO amplifier	**K**	50 Ω transmission line
E	Bias line	**L**	Waveguide-to-microstrip adaptor, RF in
F	Radial stub	**M**	Interface to frequency conversion section
G	High-impedance transmission line	**N**	Interface to IF section

Figure 1-2: A 14.4–15.35 GHz receiver module itself consisting of cascaded modules interconnected by microstrip transmission lines. Surrounding the microwave circuit are DC conditioning and control circuitry. RF in is 14.4 GHz to 15.35 GHz, LO in is 1600.625 MHz to 1741.875 MHz. The frequency of the IF is 70–1595 MHz. Detail of the frequency conversion section mounted on the mat is shown in Figure 1-3. The silk-screened mat provides a strain-relieving surface for mounting ceramic- and semiconductor-based modules.

enable maximum power transfer. The fifth volume in this series [5] considers amplifier and oscillator design and develops the skills required to develop modules.

The books in the Microwave and RF Design series are:

- Microwave and RF Design: Radio Systems
- Microwave and RF Design: Transmission Lines
- Microwave and RF Design: Networks
- Microwave and RF Design: Modules
- Microwave and RF Design: Amplifiers and Oscillators



A	Reference LO amplifier	*K*	RF amplifier
B	Power supply decoupling capacitor	*L*	Power supply decoupling capacitor
C	×2 frequency multiplier	*M*	Power supply decoupling capacitor
D	Edge-coupled PCL bandpass filter	*N*	Voltage variable attenuator
E	Microstrip bend	*O*	Power supply decoupling capacitor
F	Microstrip transmission line	*P*	Power supply decoupling capacitor
G	×2 frequency multiplier	*Q*	RF amplifier
H	Power supply decoupling capacitor	*R*	Microstrip transmission line and bend
I	stepped impedance lowpass filter	*S*	Edge-coupled PCL bandpass filter
J	Microstrip transmission line	*T*	Subharmonic mixer

Figure 1-3: Frequency conversion section of the receiver module shown in Figure 1-2. The reference LO is applied to the frequency conversion section at *c*, the RF is applied at *h* following the isolator. The IF is output at *j*.

1.4 References

[1] http://www.synergymwave.com.

[2] M. Steer, *Microwave and RF Design, Radio Systems*, 3rd ed. North Carolina State University, 2019.

[3] ——, *Microwave and RF Design, Transmission Lines*, 3rd ed. North Carolina State University, 2019.

[4] ——, *Microwave and RF Design, Networks*, 3rd ed. North Carolina State University, 2019.

[5] ——, *Microwave and RF Design, Amplifiers and Oscillators*, 3rd ed. North Carolina State University, 2019.

[6] IEEE Standard 315-1975, Graphic Symbols for Electrical and Electronics Diagrams (Including Reference Designation Letters), Adopted Sept. 1975, Reaffirmed Dec. 1993. Approved by American National Standards Institute, Jan. 1989. Approved adopted for mandatory use, Department of Defense, United States of America, Oct. 1975. Approved by Canadian Standards Institute, Oct. 1975.

[7] R. Baker, *CMOS Circuit Design, Layout, and Simulation*, 2nd ed. Wiley-Interscience, IEEE Press, 2008.

Appendix

1.A RF and Microwave Circuit Schematic Symbols

1.A.1 Element and Circuit Symbols 5
1.A.2 Sources 8
1.A.3 Diodes 8
1.A.4 Bipolar Junction Transistor 8
1.A.5 Junction Field Effect Transistor 9
1.A.6 Metal-Oxide-Semiconductor FET 9

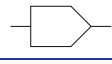
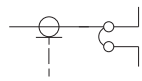
This appendix lists the symbols commonly used with RF and microwave circuits. Symbols are from IEEE Standard 315-1975 [6]. Up until the 1970s IEEE was active in establishing standard symbols for all electrical engineering fields and in particular circuit schematic symbols to be used with microwave circuits. Since then vendors of microwave computer-aided design tools have developed their own symbols but very often a vendor tends to adopt symbols similar to those used by other vendors. However there are differences and as a result there has not been a consensus to adopt a more modern standard for microwave symbols. What is presented in this chapter follows the earlier IEEE standard where possible and for components that are not in the standard, an attempt has been made to select symbols that are in common use in technical papers.

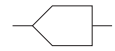
1.A.1 Element and Circuit Symbols

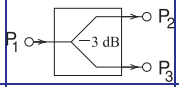
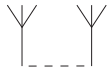
Table 1-1: IEEE standard qualifying properties added to schematic symbols to identify a particular property.

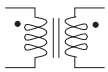
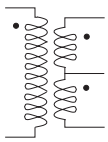
Qualifying property	Symbol
Adjustable	
Adjustable, continuously adjustable	
Adjustable, stepped	
Linear	
Nonlinear	
Positive	+
Negative	–

Table 1-2: Standard schematic symbols of RF and microwave components.

Component	Symbol	Alternate
Analog to digital converter		
Attenuator, fixed		
Attenuator, balanced		
Attenuator, unbalanced		
Attenuator, variable		
Attenuator, continuously variable		
Attenuator, stepped		
Amplifier		
Antenna, general		
Antenna, balanced		
Antenna, dipole		
Antenna, loop		
Balun		
Balun with coaxial line and dipole antenna		
Capacitor, general		
Capacitor, polarized		

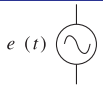
Component	Symbol	Alternate
Capacitor, variable		
Capacitor, nonlinear		
Capacitor, shielded		
Circulator		
Coaxial cable		
Conductive path		
Connector, female		
Connector, male		
Contact, fixed		
Contact, closed		
Contact, open		
Delay		
Digital-to-analog converter (DAC)		
Element, linear (* to be replaced by designation)		
Ground, general		
Ground, chassis		
Coupler		
Filter, bandpass filter (BPF)		
Filter, lowpass filter (LPF)		

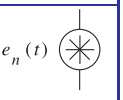
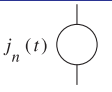
Component	Symbol	Alternate
Filter, highpass filter (HPF)		
Filter, bandstop filter (BSF)		
Isolator		
Inductor, general		
Inductor with magnetic core		
Junction		
Junction of paths		
Network, linear (* to be replaced by designation)		
Open		
Phase shifter		
Piezoelectric resonator		
Port		
Power divider		
Radio link		
Radio link with antennas		
Rectifier		
Resistor, general		
Resistor, variable		
Resistor, nonlinear		
Resistor with open path		
Resistor with short		

Component	Symbol	Alternate
Shield		
Short, movable		
Source, AC		
Source, DC		
Switch, multiposition		
Test, point		
Transformer		
Transformer with magnetic core		
Transformer, center tapped		
Triax		
Twinax		
Twinax with shield showing connection		
Twinax with shield grounded		
Short		
Wire		
Wires, connected		
Wires, unconnected, crossing		

1.A.2 Sources

Commonly used symbols for sources.

Component	Symbol
Voltage source	
Current source	
AC source	
Controlled voltage source	

Component	Symbol
Controlled current source	
Voltage noise source	
Current noise source	

1.A.3 Diodes

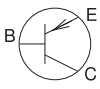
IEEE standard symbols for diodes and a rectifier [6]. ([1]In the direction of anode (A) to cathode (K). [2]Use symbol for general diode unless it is essential to show intrinsic region.)

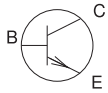
Component	Symbol
Diode, general (including Schottky)[1]	
Gunn diode[1]	
IMPATT diode[1]	
PIN diode[1,2]	
Light emitting diode (LED)[1]	

Component	Symbol
Rectifier[1]	
Tunnel diode1	
Varactor diode[1]	
Zener diode[1]	

1.A.4 Bipolar Junction Transistor

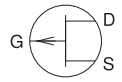
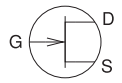
IEEE standard schematic symbols for bipolar junction transistors (BJT and HBT) [6] and commonly used symbols in layouts [7]. The letters indicate terminals: B (base), C (collector), E (emitter).

Transistor	IEEE symbol	Commonly used symbol
BJT, pnp		

Transistor	IEEE symbol	Commonly used symbol
BJT, npn		

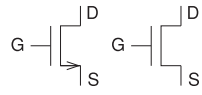
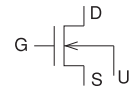
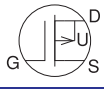
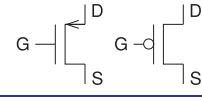
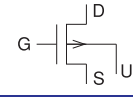
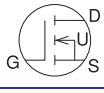
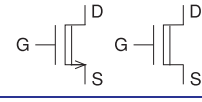
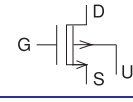
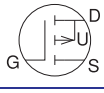
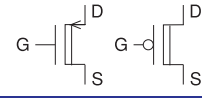
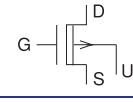
1.A.5 Junction Field Effect Transistor

Table 1-3: IEEE standard schematic symbols for junction field effect transistors (MESFET, HEMT, JFET) [6] and symbols more commonly used in schematics. The letters indicate terminals: G (gate), D (drain), S (source).

Transistor	IEEE symbol	Commonly used symbol
FET, pJFET		
FET, nJFET, MESFET, HEMT		

1.A.6 Metal-Oxide-Semiconductor Field Effect Transistor

Table 1-4: IEEE standard schematic symbols for MOSFET transistors [6] and symbols more commonly used in schematics [7]. The MOSFET symbols are for enhancement- and depletion-mode transistors. The letters indicate terminals: G (gate), D (drain), S (source), U (bulk). Four-terminal and three-terminal common symbols are shown. The three-terminal common symbol is most often used when the bulk is connected to the most negative connection in the circuit, and the pMOSFET symbol is used when the bulk is tied to V_{DD} (the most positive connection). The bulk connection is often not shown, as it is assumed to be connected to the most negative voltage point.

Transistor	IEEE symbol	Commonly used symbol (three-terminal)	Commonly used symbol (four-terminal)
FET, nMOS, depletion			
FET, pMOS, depletion			
FET, nMOS, enhancement			
FET, pMOS, enhancement			

CHAPTER 2

Filters

With Wael Fathelbab

2.1 Introduction 11
2.2 Singly and Doubly Terminated Networks 13
2.3 The Lowpass Filter Prototype 17
2.4 The Maximally Flat (Butterworth) Lowpass Approximation 18
2.5 The Chebyshev Lowpass Approximation 22
2.6 Element Extraction 24
2.7 Butterworth and Chebyshev Filters 29
2.8 Impedance and Admittance Inverters 33
2.9 Filter Transformations 40
2.10 Cascaded Line Realization of Filters 46
2.11 Butterworth and Chebyshev Bandpass Filters 48
2.12 Richards's Transformation 50
2.13 Kuroda's and Norton's Network Identities 54
2.14 Inter-resonator Coupled Filters 57
2.15 Bandpass Filter Topologies 61
2.16 Design of a Bandstop Filter 62
2.17 Active Filters 64
2.18 Transient Response of a Bandpass Filter 69
2.19 Summary 71
2.20 References 72
2.21 Exercises 74

2.1 Introduction

Filters are the most fundamental of signal processing circuits using energy storage elements to obtain frequency-dependent characteristics. Some of the important filter attributes are (1) controlling noise by not allowing out-of-band noise to propagate in a circuit; (2) keeping signals outside the transmit band, especially harmonics, from being transmitted; and (3) presenting only signals in a specified band to active receive circuitry. At microwave frequencies a filter can consist solely of lumped elements, solely of distributed elements, or a mix of lumped and distributed elements. The distributed realizations can be transmission line-based implementations of

the components of lumped-element filter prototypes or, preferably, make use of the particular frequency characteristics found with certain distributed structures. Loss in lumped elements, particularly above a few gigahertz, means that the performance of distributed filters nearly always exceeds that of lumped-element filters. However, since the basic component of a distributed filter is a one-quarter wavelength long transmission line, distributed filters can be prohibitively large below a few gigahertz.

Only a few basic types of responses are required of most RF filters as follows:

(a) Lowpass—providing maximum power transfer at frequencies below the corner frequency, f_0. Above f_0, transmission is blocked. See Figure 2-1(a).

(b) Highpass—passing signals at frequencies above f_0. Below f_0, transmission is blocked. See Figure 2-1(b).

(c) Bandpass—passing signals at frequencies between lower and upper corner frequencies (defining the passband) and blocking transmission outside the band. This is the most common type of RF filter. See Figure 2-1(c).

(d) Bandstop (or notch)—which blocks signals between lower and upper corner frequencies (defining the stopband). See Figure 2-1(d).

(e) Allpass—which equalizes a signal by adjusting the phase generally to correct for phase distortion elsewhere. See Figure 2-1(e).

The above list is not comprehensive, as actual operating conditions may mandate specific frequency profiles. For example, WiFi systems operating at 2.45 GHz are susceptible to potentially large signals being transmitted from nearby cellular phones operating in the 1700–2300 MHz range. Thus a front-end filter for a 2.45 GHz system must ensure very high levels of attenuation over the 1700–2300 MHz range. Thus the optimum solution here, probably, will have an asymmetrical frequency response with high rejection on one side of the passband obtained by accepting lower rejection on the other side.

2.1.1 Filter Prototypes

In the 1960s an approach to RF filter design and synthesis was developed and this is still followed. The approach is to translate the mathematical response of a lowpass filter. A filter with the desired lowpass response is then synthesized using lumped elements and the resulting filter is called a lowpass prototype. The lowpass filter is then transformed so that the new lumped-element filter has the desired RF response, such as highpass or bandpass. In the case of a bandpass filter, each inductor and capacitor of the lowpass prototype becomes a resonator that is coupled to another resonator. In distributed form, this basic resonator is a one-quarter wavelength long transmission line.

Filter synthesis is a systematic approach to realizing circuits with desired frequency characteristics. A filter can also be thought of as a hardware implementation of a differential equation producing a specific impulse,

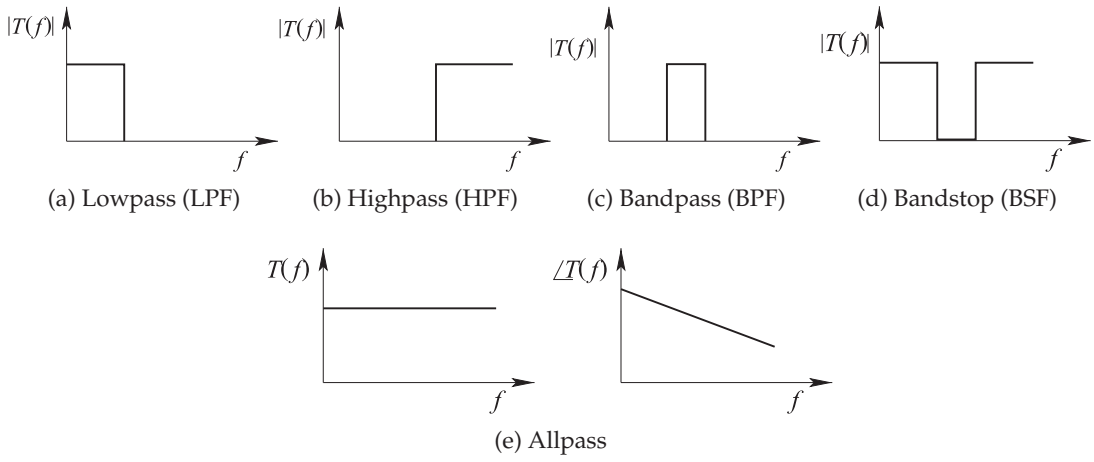


Figure 2-1: Ideal filter transfer function, $T(f)$, responses.

step, or frequency response. It is not surprising that a filter can be described this way as the lumped reactive elements, L and C, describe first-order differential equations and their interconnection describes higher-order differential equations.

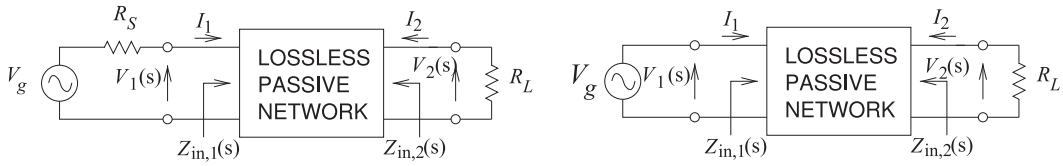
The synthesis of a filter response begins by identifying the desired differential equation response. The differential equations are specified using the Laplace variable, s, so that in the frequency domain (with $s = \jmath\omega$) an nth-order differential equation becomes an nth-order polynomial expression in s. If possible, and this is usually the case, a lowpass version of a filter can be developed. The conversion to the final filter shape, say a bandpass response, proceeds mathematically through a number of stages. Each stage has a circuit form and each stage is called a filter prototype.

2.1.2 Image Parameter Versus Insertion loss Methods

Filter networks can be synthesized using the image parameter method or the reflection coefficient (or insertion loss) method [1]. The image parameter measured is based on cascading two-port networks with each two-port having the essential desired filter response [2]. Several of these two-ports are cascaded to achieve a periodic-like structure that has the final desired filter response. This method is restrictive and arbitrary filter responses can not be obtained. It is rarely used in RF and microwave filter design but is useful in analyzing simple structures with cascades of identical elements [3]. A better approach is to use the insertion loss method as used in this chapter.

2.2 Singly and Doubly Terminated Networks

Filters are generally two-port networks and they provide maximum power transfer from a source to a load over a specified frequency range while rejecting the transmission of signals at other frequencies. Two possible filter networks are shown in Figure 2-2. The network in Figure 2-2(a) is known as a doubly terminated network, as both ports are resistively terminated.



(a) Doubly terminated network (b) Singly terminated network

Figure 2-2: Terminated networks.

The network in Figure 2-2(b) is called a singly terminated network, as only one port is terminated in a resistor. The doubly terminated network is much closer to the type of network required at RF where loads and source impedances are finite. A singly terminated network is applicable in some RF integrated circuit applications where very little RF power is involved and often when feedback is used. In such cases the output of an RFIC amplifier can approximate an ideal voltage source as the Thevenin equivalent source impedance can be negligible. Most synthesis of singly terminated filter networks uses an analogous procedure to that presented here for doubly terminated filters.

2.2.1 Doubly Terminated Networks

The established filter synthesis procedure for doubly terminated filter networks focuses on realizing the input reflection coefficient. For a lossless filter the square of the reflection coefficient magnitude is one minus the square of the insertion loss and this is the origin of this method being called the insertion loss method.

The input reflection coefficient of the doubly terminated network in Figure 2-2(a) is

$$\Gamma_1(s) = \frac{Z_{\text{in},1}(s) - R_S}{Z_{\text{in},1}(s) + R_S}, \tag{2.1}$$

where the reference impedance is the source resistance, R_S, and s is the Laplace variable. In the passband of a filter the reflection coefficient is approximately zero.

There are several other parameters used in filter design and these will be introduced now. The **transducer power ratio** (**TPR**) is defined as

$$\begin{aligned} \text{TPR} &= \frac{\text{Maximum power available from the source}}{\text{Power absorbed by the load}} \\ &= \frac{\frac{1}{2}(V_g(s)/2)^2/R_S}{\frac{1}{2}V_2^2(s)/R_L} = \left|\frac{1}{2}\sqrt{\frac{R_L}{R_S}}\frac{V_g(s)}{V_2(s)}\right|^2. \end{aligned} \tag{2.2}$$

The transmission coefficient, $T(s)$, of the network is

$$T(s) = \frac{1}{\sqrt{\text{TPR}(s)}} = 2\sqrt{\frac{R_S}{R_L}}\frac{V_2(s)}{V_g(s)}, \tag{2.3}$$

IF the filter is lossless, the insertion loss (IL) (or **transducer function**) is

$$\text{IL}(s) = \text{TPR}(s) = |1/T(s)|^2 . \tag{2.4}$$

The next step in development of the filter synthesis procedure is introduction of the **characteristic function**, defined as the ratio of the reflection and transmission coefficients:

$$K(s) = \frac{\Gamma_1(s)}{T(s)} = \frac{N(s)}{D(s)}, \tag{2.5}$$

where N is the numerator function and D is the denominator function. Ideally the filter network is lossless and so, from the principle of energy conservation, the sum of the magnitudes squared of the transmission and reflection coefficients must be unity:

$$|T(s)|^2 + |\Gamma_1(s)|^2 = 1. \tag{2.6}$$

That is,

$$|T(s)|^2 = 1 - |\Gamma_1(s)|^2. \tag{2.7}$$

Dividing both sides of Equation (2.7) by $|T(s)|^2$ results in

$$1 = \frac{1}{|T(s)|^2} - \frac{|\Gamma_1(s)|^2}{|T(s)|^2}, \tag{2.8}$$

or

$$1 = |\mathrm{IL}(s)| - |K(s)|^2. \tag{2.9}$$

Rearranging Equation (2.9) leads to

$$|\mathrm{IL}(s)| = 1 + |K(s)|^2 \tag{2.10}$$

and so

$$|T(s)|^2 = \frac{1}{1 + |K(s)|^2} \tag{2.11}$$

and

$$|\Gamma_1(s)|^2 = \frac{|K(s)|^2}{1 + |K(s)|^2}. \tag{2.12}$$

In the above it is seen that both the reflection and transmission coefficients are functions of the characteristic function of the two-port network. A milestone has been reached. In an RF filter, the frequency-dependent insertion loss or transmission coefficient is of most importance, as these are directly related to power flow. Equation (2.11) shows that this can be expressed in terms of another function, $K(s)$, which, from Equation (2.5), can be expressed as the ratio of $N(s)$ and $D(s)$. For lumped-element circuits, $N(s)$ and $D(s)$ are polynomials.

2.2.2 Lowpass Filter Response

As an example of the relationship of the various filter responses, consider the lossless lowpass filter responses shown in Figure 2-3. This filter has a lowpass response with a corner frequency expressed in radians as $\omega_c = 1$ rad/s. Ideally $T(s)$ for $s \leq \jmath$ (note $s = \jmath\omega$) would be one, and for $s > \jmath$, $T(s)$ would be zero. It is not possible to realize such an ideal response, and the response shown in Figure 2-3(a) is typical of what can be achieved. The reflection coefficient is shown in Figure 2-3(b), with the characteristic function $K(s)$ shown in Figure 2-3(c). If $|K(s)|^2$ is expressed as the ratio of two polynomials (i.e, as $N(s)/D(s)$), then it can be seen from Figure 2-3(e and f) that the zeros of the reflection coefficient, $|\Gamma_1(s)|^2$, are also the zeros of $N(s)$. Also, it is observed that the zeros of the transmission coefficient are also the zeros of $D(s)$, as shown in Figure 2-3(a and c).

The first objective in lumped-element filter design is development of the transfer function of the network in the frequency domain or, equivalently, the s domain.[1] The input-output transfer function of the generic filter in Figure 2-4 is

[1] Care is being taken in the use of terminology as s can be complex with some filter types, but these filters are realized using digital signal processing.

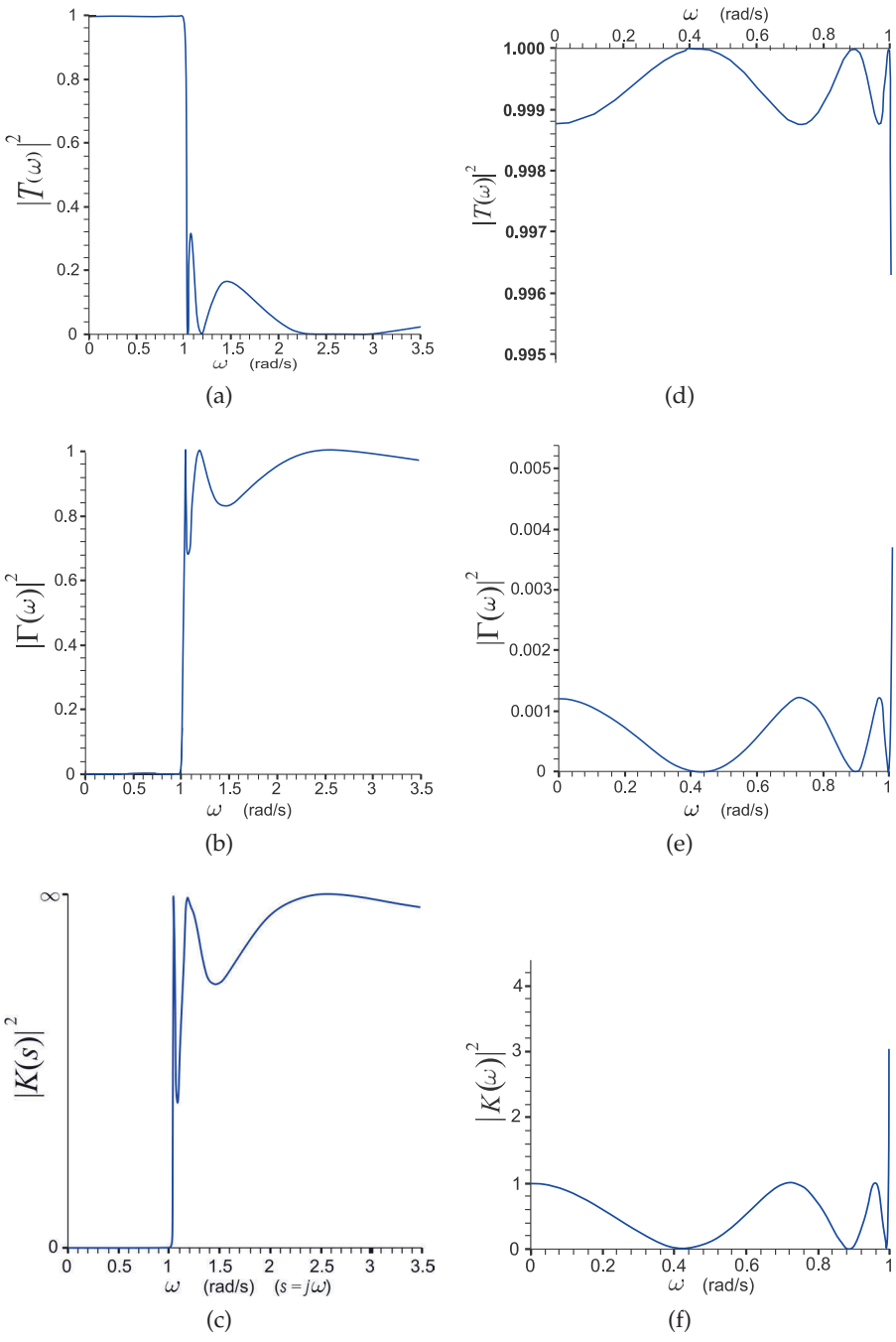


Figure 2-3: An example of a lowpass filter in terms of various responses: (a) transmission coefficient; (b) reflection coefficient response; and (c) characteristic function response. Detailed responses are shown in (d), (e), and (f), respectively.

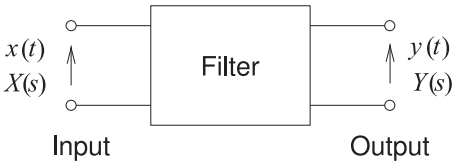


Figure 2-4: Generic filter.

$$T(s) = Y(s)/X(s) \tag{2.13}$$

and the design procedure is to match this response to a response defined by the ratio of two polynomials:

$$T(s) = \frac{N(s)}{D(s)} = \frac{a_m s^m + a_{m-1}s^{m-1} + \cdots + a_1 s + a_0}{s^n + b_{n-1}s^{n-1} + \cdots + b_1 s + b_0}. \tag{2.14}$$

Here N stands for numerator and D for denominator, and these are not the same as those in Equation (2.5) (where they were just labels for numerator and denominator). The filter response using a **pole-zero description** can be synthesized so the design process begins by rewriting Equation (2.14) explicitly in terms of zeros, z_m, and poles, p_n:

$$T(s) = \frac{N(s)}{D(s)} = \frac{a_m(s+z_1)(s+z_2)\cdots(s+z_{m-1})(s+z_m)}{(s+p_1)(s+p_2)\cdots(s+p_{n-1})(s+p_n)}. \tag{2.15}$$

Since only the frequency response is of interest, $s = \jmath\omega$, thus simplifying the analysis for sinusoidal signals.

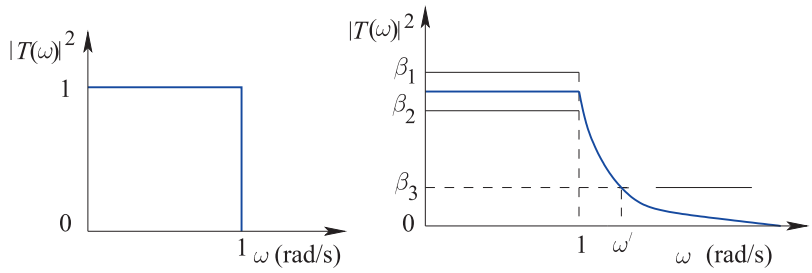
The poles and zeros can be complex numbers and can be plotted on the complex s plane. Conditions imposed by realizable circuits require that $D(s)$ be a **Hurwitz** polynomial,[2] which ensures that its poles are located in the left-half plane. $N(s)$ determines the location of the transmission zeros of the filter, and the order of $N(s)$ cannot be more than the order of $D(s)$. That is, $n \geq m$ so that the filter has finite or zero response at infinite frequency.

Two strategies can be employed in deriving the filter response. The first is to derive the polynomials $N(s)$ and $D(s)$ in Equation (2.14). This seems like an open-ended problem, but it was discovered in the 1950s and 1960s that in normal situations there are only a few types of useful responses that are described by a few polynomials, including Butterworth, Bessel, Chebyshev, and Cauer polynomials.

2.3 The Lowpass Filter Prototype

Most filter design is based on the synthesis of a lowpass filter equivalent called a lowpass prototype. Transformations are then used to correct for the actual source and load impedances, frequency range, and the desired filter type, such as highpass or bandpass. An ideal lowpass response is shown in Figure 2-5(a), which shows a transmission coefficient of 1 up to a normalized frequency of 1 rad/sec. This type of response defines what is called a **brick-wall filter**, which unfortunately cannot be physically realized. Instead, the response is approximated and specified in terms of a response template (see Figure 2-5(b)).

[2] A Hurwitz polynomial is a polynomial whose coefficients are positive real numbers and whose zeros are located in the left-half of the complex s plane.



(a) Ideal brick-wall response (b) Response template

Figure 2-5: Lowpass filter response.

The template specifies a passband response that is between β_1 and β_2 at frequencies below the corner radian frequency (at $\omega = 1$), and below β_3 above a radian frequency $\omega' > 1$. The lowpass filter is harder to realize the closer the specified response is to the ideal response shown in Figure 2-5(a), that is, when

$$(\beta_2 - \beta_1) \to 0, \quad \beta_3 \to 0, \quad \text{and} \quad (\omega' - 1) \to 0. \tag{2.16}$$

Several polynomials have particularly interesting characteristics that match the requirements of the response template. At first it may seem surprising that polynomial functions could yield close to ideal filter responses. However, as will be seen, there is something special about these polynomials: they are natural solutions to extreme conditions. For example, the nth-order Butterworth polynomial has the special property that the first n derivatives at $s = 0$ are zero. The other polynomials also have extreme properties, and filters that are synthesized using them also have extreme properties. As well as the maximally flat property resulting from Butterworth polynomials, filter responses obtained using Chebyshev polynomials have the steepest skirts (i.e., fastest transitions from the frequency region where signals are passed to the region where they are blocked). Usually it is one of these extreme cases that is most desirable.

A streamlining of the filter synthesis procedure is obtained by first focusing on the development of normalized lowpass filters having a 1 rad/s corner frequency and 1 Ω reference impedance. Transformations transform a lowpass filter prototype into another filter having the response desired. These transformations give rise to symmetrical responses.

2.4 The Maximally Flat (Butterworth) Lowpass Approximation

Butterworth filters have a maximally flat response (Figure 2-6) that in the time domain corresponds to a critically damped system.

2.4.1 Butterworth Filter Design

Butterworth filters have the transfer function

$$T(s) = \frac{N(s)}{D(s)} = \frac{k}{s^n + b_{n-1}s^{n-1} + \cdots + b_1 s + b_0}. \tag{2.17}$$

This is an all-pole response. The characteristic polynomial of the Butterworth filter is

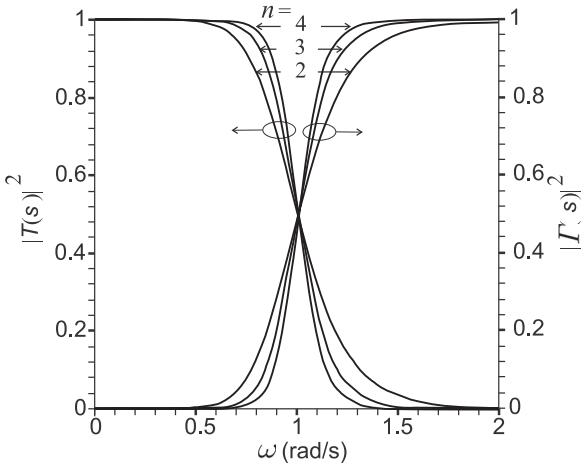


Figure 2-6: Maximally flat, or Butterworth, low-pass filter approximation for various orders, n, of the filter.

n	Factors of $B_n(s)$
1	$(s+1)$
2	$(s^2+1.4142s+1)$
3	$(s+1)(s^2+s+1)$
4	$(s^2+0.7654s+1)(s^2+1.8478s+1)$
5	$(s+1)(s^2+0.6180s+1)(s^2+1.6180s+1)$
6	$(s^2+0.5176s+1)(s^2+1.4142s+1)(s^2+1.9319s+1)$
7	$(s+1)(s^2+0.4450s+1)(s^2+1.2470s+1)(s^2+1.8019s+1)$
8	$(s^2+0.3902+1)(s^2+1.1111s+1)(s^2+1.6629s+1)(s^2+1.9616s+1)$

Table 2-1: Factors of the $B_n(s)$ polynomial.

$$|K(s)|^2 = \left|s^{2n}\right| = \omega^{2n}, \tag{2.18}$$

since $s = \jmath\omega$ and where n is the order of the function. Thus the transmission coefficient is

$$|T(s)|^2 = \frac{1}{1+|K(s)|^2} = \frac{1}{1+|s^{2n}|} = \frac{1}{1+\omega^{2n}}, \tag{2.19}$$

which is often written as

$$|T(s)|^2 = \frac{1}{1+|K(s)|^2} = \frac{1}{B_n(s)B_n(-s)}. \tag{2.20}$$

In the filter community, $B_n(s)$ is called the Butterworth polynomial and has the general form

$$B_n(s) = \begin{cases} \prod_{k=1}^{n/2}\left[s^2 - 2s\cos\left(\dfrac{2k+n-1}{2n}\pi\right)+1\right] & \text{for } n \text{ even} \\ (s+1)\prod_{k=1}^{(n-1)/2}\left[s^2 - 2s\cos\left(\dfrac{2k+n-1}{2n}\pi\right)+1\right] & \text{for } n \text{ odd} \end{cases}. \tag{2.21}$$

$B_n(s)$ is given in factorized form in Table 2-1. Further factorization of the second-order factors results in complex conjugate roots, so they are generally left in the form shown.

Insight is gained by examining the characteristic polynomial for a real radian frequency, ω (where $s = \jmath\omega$), at the following frequency points:

$$\text{at } \omega = 0 \quad \rightarrow \quad |T(0)|^2 = \frac{1}{1+0} = 1 \tag{2.22}$$

$$\text{at } \omega = 1 \quad \rightarrow \quad |T(1)|^2 = \frac{1}{1+1} = \frac{1}{2}. \tag{2.23}$$

Thus the transmission response of the filter is at half power at the corner frequency $\omega = 1$. Another observation is that the transmission coefficient at $\omega = 0$ and $\omega = 1$ are independent of the degree of the characteristic polynomial. This is shown in Figure 2-6, where the responses of Butterworth filters are plotted for three different orders.

2.4.2 Construction of the Transfer Function

Often the characteristic function, K, is specified in terms of the variable ω, and for synthesis purposes the response must be transformed back to the s domain. As an example, consider the maximally flat function of third order with a characteristic polynomial equal to ω^3 (from Equation (2.19)):

$$K(\omega) = \omega^3 \quad \text{and} \quad |T(\omega)|^2 = \frac{1}{1+\omega^6}. \tag{2.24}$$

(Note that here the first six derivatives of $|K(s)|^2$ with respect to s are zero at $s = 0$.) So, from Equation (2.7),

$$|\Gamma_1(\omega)|^2 = 1 - \frac{1}{1+\omega^6} = \frac{\omega^6}{1+\omega^6}. \tag{2.25}$$

At real frequencies (i.e., frequencies lying on the imaginary axis in the s plane)

$$s = \jmath\omega, \quad \text{so} \quad \omega = s/\jmath = -\jmath s. \tag{2.26}$$

Thus, in the s domain, the reflection coefficient becomes

$$|\Gamma_1(\jmath\omega)|^2 = |\Gamma_1(\jmath(-\jmath s))|^2 = |\Gamma_1(s)|^2 = \frac{(-\jmath s)^6}{1+(-\jmath s)^6}. \tag{2.27}$$

That is, $$|\Gamma_1(s)|^2 = \frac{-s^6}{1-s^6} = \Gamma_1(s)\,\Gamma_1(-s). \tag{2.28}$$

Factoring the denominator polynomial yields

$$1 - s^6 = (1-s)(1+s)\left(s^2+s+1\right)\left(s^2-s+1\right). \tag{2.29}$$

By choosing those factors with roots only in the left-half plane (required for a realizable network), the zeros of the denominator are obtained. Now the numerator of $|\Gamma_1(s)|^2$ is easily factored, and since this is third order, it only has three reflection zeros at DC. The resulting function is

$$\Gamma_1(s) = \frac{s^3}{(s+1)\left(s^2+s+1\right)}, \tag{2.30}$$

and so $$T(s) = \frac{1}{(s+1)(s^2+s+1)}. \tag{2.31}$$

This example illustrates how $\Gamma_1(s)$ and $T(s)$ can be obtained from $K(s)$. This procedure is generalized in the next section.

2.4.3 nth-Order Reflection Approximation

Following the procedure outlined in the previous section, generalization leads to the nth-order Butterworth response:

$$|\Gamma_1(s)|^2 = \Gamma_1(s)\cdot\Gamma_1(-s) = \frac{(-s^2)^n}{1+(-s^2)^n} = \frac{(-s^2)^n}{\prod_{i=1}^{n}(s-s_i)\cdot\prod_{j=n+1}^{2n}(s-s_j)}. \tag{2.32}$$

So $|\Gamma_1(s)|^2$ has n roots (these are the s_is) lying in the left-half s plane, and n roots (the s_js) lying in the right-half s plane. Note that j is used as an index and $\jmath$ ($\jmath$, without the dot) represents $\sqrt{-1}$. It is reasonable to group all of the left-half plane roots together (a circuit could not be synthesized if there were a right-half plane root) so that

$$\Gamma_1(s) = \frac{(-s)^n}{\prod_{i=1}^{n}(s-s_i)}. \tag{2.33}$$

Solving for the roots of the denominator of Equation (2.32) (i.e., finding the roots of $1+\left(-s^2\right)^n = 0$) yields the following roots in the left-half s plane:

$$s_i = \exp\left\{\jmath\,(2i-1+n)\frac{\pi}{2n}\right\} \quad i = 1, 2, ..., n, \tag{2.34}$$

where $\exp(\jmath\theta) = \cos(\theta) + \jmath\sin(\theta)$.

EXAMPLE 2.1 Reflection Coefficient Derivation for Butterworth Filter

Develop the reflection coefficient from the roots of the third-order ($n = 3$) Butterworth lowpass filter prototype.

Solution: From Equation (2.34) the three roots are

$$s_1 = \exp\left\{\jmath\,(2\times1-1+3)\frac{\pi}{2\times3}\right\} = \exp\left\{\jmath\,\frac{2}{3}\pi\right\} \tag{2.35}$$

$$s_2 = \exp\left\{\jmath\,(2\times2-1+3)\frac{\pi}{2\times3}\right\} = \exp\{\jmath\,\pi\} \tag{2.36}$$

$$s_3 = \exp\left\{\jmath\,(2\times3-1+3)\frac{\pi}{2\times3}\right\} = \exp\left\{\jmath\,\frac{4}{3}\pi\right\} \tag{2.37}$$

and the input reflection coefficient of the filter normalized to 1 Ω is

$$\Gamma_1(s) = \frac{s^3}{\left(s+\frac{1}{2}-\jmath\frac{\sqrt{3}}{2}\right)(s+1)\left(s+\frac{1}{2}+\jmath\frac{\sqrt{3}}{2}\right)} = \frac{s^3}{(s+1)(s^2+s+1)}. \tag{2.38}$$

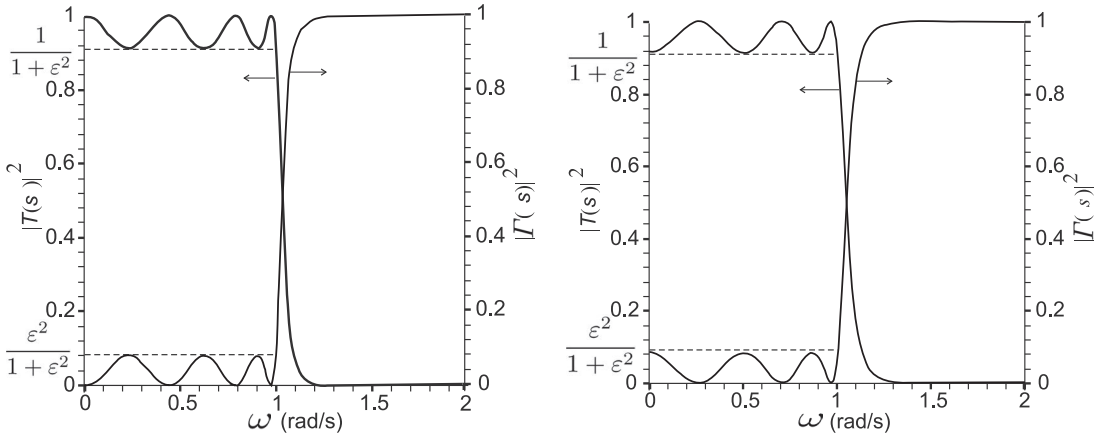
This is the same as Equation (2.30).

2.4.4 Bandwidth Consideration

At the corner frequency of the Butterworth response, that is, at 1 rad/s, the transmission response is 3 dB down from its maximum transfer response.

Table 2-2: Radian frequencies at which the response of a Butterworth filter is 1 dB down for a corner frequency $f_0 = 0.1592$ Hz ($\omega_0 = 1$ rad/s). At f_0 the transmission response is down 3 dB.

Order, n							
$n = 3$	4	5	6	7	8	9	10
0.798 rad/s	0.844	0.873	0.901	0.908	0.919	0.928	0.935



(a) Seventh-order filter, $n = 7$ (b) Sixth-order filter, $n = 6$

Figure 2-7: Chebyshev lowpass filter responses. In decibels, the ripple is $R_{\text{dB}} = -10\log[1/(1+\varepsilon^2)] = 10\log(1+\varepsilon^2)$, ε is called the ripple factor.

Sometimes microwave filters are designed to have 1 dB bandwidths. The radian frequencies at which the responses of various orders of Butterworth filters are 1 dB down are given in Table 2-2. By frequency scaling the Butterworth response the filter can be design for a specified 1 dB bandwidth.

2.5 The Chebyshev Lowpass Approximation

The maximally flat approximation to the ideal lowpass filter response is best near the origin but not so good near the band edge. Chebyshev filters have better responses near the band edge, with lower insertion loss near the edges, but at the cost of ripples in the passband. Example reflection and transmission responses are shown in Figure 2-7 for a seventh-order and a sixth-order Chebyshev lowpass filter.

2.5.1 Chebyshev Filter Design

The general form of the Chebyshev transmission coefficient is

$$|T(s)|^2 = \frac{1}{1+\varepsilon^2|K(s)|^2}, \tag{2.39}$$

where ε is the ripple factor and defines the **passband ripple** (**PBR**):

$$\text{PBR} = 1+\varepsilon^2, \quad \text{or in decibels} \quad R_{\text{dB}} = \text{PBR}\,|_{\text{dB}} = 10\log\left(1+\varepsilon^2\right). \tag{2.40}$$

The PBR can be seen in the transmission response, $|T(s)|^2$, in Figure 2-7. In the passband the peaks of the lossless filter response have $|T(s)|^2 = 1$ and the minimums of the ripple response all have $|T(s)|^2 = 1/\left(1+\varepsilon^2\right) = 1/\text{PBR}$. Consequently Chebyshev filters are also known as **equiripple** all-pole lowpass filters. Also note that the corner radian frequency, $\omega = 1$ for the lowpass filter prototype, has a transmission response (i.e., insertion loss IL) of $|T(s)|^2 = 1/\left(1+\varepsilon^2\right)$, whereas the Butterworth transmission response was at half power at the corner frequency. For the Chebyshev filter, the insertion loss at the corner frequency is the ripple:

$$\text{IL} = 1R_{\text{dB}} = 10\log(1+\varepsilon^2). \tag{2.41}$$

For the nth-order Chebyshev (lowpass filter) approximation, the square of the characteristic function is

$$|K_n(\omega)|^2 = \begin{cases} \cos^2\left[n\cos^{-1}(\omega)\right], & -1 \le \omega \le 1 \\ \cosh^2\left[n\cosh^{-1}(|\omega|)\right], & \omega \le -1,\ \omega \ge 1 \end{cases}, \tag{2.42}$$

which can be expressed as a polynomial. For example, with $n = 3$,

$$K_3(\omega) = 4\omega^3 - 3\omega, \quad \text{for all } \omega. \tag{2.43}$$

(This equivalence was derived by Pafnuty Chebyshev.) It is surprising that the trigonometric expression has such a simple polynomial equivalence. From Equation (2.11) the transmission coefficient is (for $-1 \le \omega \le 1$)

$$|T(\omega)|^2 = \frac{1}{1+\varepsilon^2\cos^2\left[n\cos^{-1}(\omega)\right]} \tag{2.44}$$

and the reflection coefficient is

$$|\Gamma_1(\omega)|^2 = \frac{\varepsilon^2\cos^2\left[n\cos^{-1}(\omega)\right]}{1+\varepsilon^2\cos^2\left[n\cos^{-1}(\omega)\right]}. \tag{2.45}$$

Factorizing the denominator of either Equation (2.44) or Equation (2.45) yields the following roots (of the denominators of $\Gamma_1(s)$ and $T(s)$):

$$\begin{aligned} s_i = &\sin\left[\frac{(2i-1)\pi}{2n}\right]\sinh\left[\frac{1}{n}\sinh^{-1}\left(\frac{1}{\varepsilon}\right)\right] \\ &+ \jmath\cos\left[\frac{(2i-1)\pi}{2n}\right]\cosh\left[\frac{1}{n}\sinh^{-1}\left(\frac{1}{\varepsilon}\right)\right] \qquad i = 1,2,\ldots,n. \end{aligned} \tag{2.46}$$

The roots of the numerator of $\Gamma_1(s)$ in the s plane are

$$s_k = \jmath\cos\frac{(2k-1)\pi}{2n} \qquad k = 1,2,\ldots,n. \tag{2.47}$$

Equations (2.46) and (2.47) can be used to obtain the reflection and transmission coefficients directly in the s domain.

2.5.2 Chebyshev Approximation and Recursion

The characteristic function of the Chebyshev approximation can be obtained from the recursion formula,

$$K_n(\omega) = 2\omega K_{n-1}(\omega) - K_{n-2}(\omega), \tag{2.48}$$

Table 2-3: Radian frequencies at which the transmission response of an nth Chebyshev filter is down 1 dB and 3 dB for a corner frequency $\omega_0 = 1$ rad/s. (Note that ω_0 is the radian frequency at which the transmission response of a Chebyshev filter is down by the ripple, see Figure 2-7.)

Response 1 dB down				
Ripple	$n = 3$	$n = 5$	$n = 7$	$n = 9$
0.01 dB	1.564	1.192	1.097	1.058
0.1 dB	1.202	1.071	1.036	1.022
0.2 dB	1.127	1.045	1.023	1.014
1 dB	1.000	1.000	1.000	1.000
3 dB	–	–	–	–

Response 3 dB down				
Ripple	$n = 3$	$n = 5$	$n = 7$	$n = 9$
0.01 dB	1.877	1.291	1.145	1.087
0.1 dB	1.389	1.134	1.068	1.041
0.2 dB	1.284	1.099	1.050	1.030
1 dB	1.095	1.0338	1.017	1.010
3 dB	1.000	1.000	1.000	1.000

$$\text{with} \qquad K_1(\omega) = \omega; \qquad K_2(\omega) = 2\omega^2 - 1. \tag{2.49}$$

For example, with $n = 3$,

$$\begin{aligned} K_3(\omega) &= 2\omega K_{3-1}(\omega) - K_{3-2}(\omega) \\ &= 2\omega\left(2\omega^2 - 1\right) - \omega = 4\omega^3 - 2\omega - \omega = 4\omega^3 - 3\omega. \end{aligned} \tag{2.50}$$

2.5.3 Bandwidth Consideration

At the corner frequency of the Chebyshev filter the transmission response is down by the amount of the ripple. This can be seen in Figure 2-7. However, the bandwidth of a filter is usually specified in terms of its 1 dB or 3 dB bandwidth at which the transmission response is down 1 dB or 3 dB, respectively, from its maximum response. The radian frequencies at which the responses of various orders of Chebyshev filters are 1 dB down and 3 dB down are given in Table 2-3. By frequency scaling the Chebyshev response, the filter can be designed for a specified 1 dB or 3 dB bandwidth.

2.6 Element Extraction

In the previous two sections the mathematical responses of Butterworth and Chebyshev filters were derived for various orders. In this section it will be shown how these filters can be implemented with inductors and capacitors using what is called **ladder synthesis** [4].

2.6.1 Ladder Synthesis

To obtain the element values yielding the desired transfer function, an impedance or admittance function must first be obtained. The impedance or admittance function can be readily obtained from the input reflection coefficient of a network, but for now the focus is on synthesizing a given impedance function. A general impedance function can be expressed as

$$Z(s) = \frac{a_n\left(s^2 + \omega_1^2\right)\left(s^2 + \omega_3^2\right)\left(s^2 + \omega_5^2\right)\ldots}{b_m s\left(s^2 + \omega_2^2\right)\left(s^2 + \omega_4^2\right)\left(s^2 + \omega_6^2\right)\ldots}, \tag{2.51}$$

where a_n and b_m are constants. This can be realized using L and C elements in a network terminated by a resistor, provided that the degree of the numerator and denominator differ by no more than unity (i.e., $|m-n| \leq 1$). In

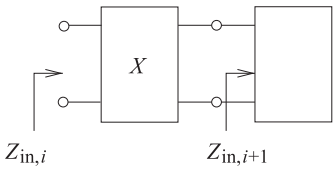


Figure 2-8: Extraction of a network X to reduce an impedance $Z_{\text{in},i}$ to a lower-order impedance $Z_{\text{in},i+1}$.

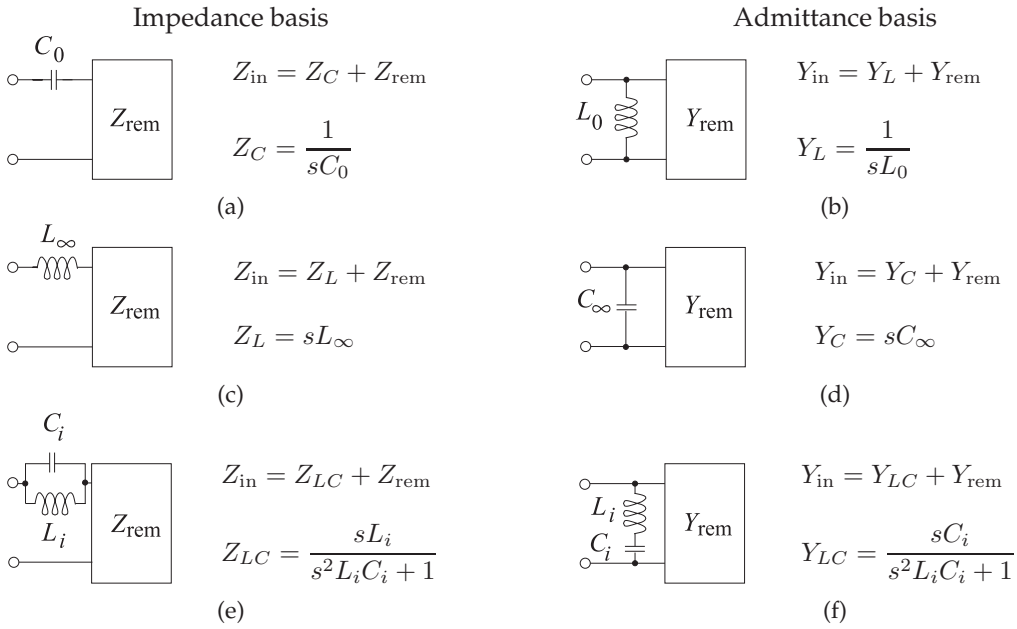


Figure 2-9: Synthesis of impedance and admittance functions. Starting with an impedance function $Z(s)$: (a) extraction of a series capacitor; (c) extraction of a series inductor; and (e) extraction of a series parallel LC block i. Starting with an admittance function $Y(s)$: (b) extraction of a shunt inductor; (d) extraction of a shunt capacitor; and (f) extraction of a shunt series LC block.

the case of a doubly terminated network, this resistor is the load. The element extraction procedure, shown in Figure 2-8, involves extracting a network X from $Z_{\text{in},i}$, leaving a reduced-order impedance $Z_{\text{in},i+1}$.

The extraction of inductors and capacitors is illustrated in Figure 2-9. Thus, following the extraction of an element or a pair of elements an impedance, Z_{rem}, or admittance, Y_{rem}, remains that can be similarly simplified. For example, and referring to Figure 2-9(a), $Z(s) = 1/(sC) + Z_{\text{rem}}$. So a pole of $Z(s)$ at DC requires the extraction of a series capacitor of value (see Figure 2-9(a))

$$C_0 = \left.\frac{1}{sZ(s)}\right|_{s=0}, \tag{2.52}$$

while a pole at infinity requires the extraction of a series inductor of value (see Figure 2-9(c))

$$L_\infty = \left.\frac{Z(s)}{s}\right|_{s=\infty}. \tag{2.53}$$

Another possibility is a pole at a finite frequency (call this ω_0), which requires the extraction of a series parallel LC block, as shown in Figure 2-9(e), with

elements of value

$$C_i = \left.\frac{s}{(s^2+\omega_0^2)\,Z(s)}\right|_{s=j\omega_o} \quad \text{and} \quad L_i = \frac{1}{\omega_0^2.C_i}. \tag{2.54}$$

The extraction process can also be carried out on an admittance basis. First,

$$Y(s) = \frac{b_m s\left(s^2+\omega_2^2\right)\left(s^2+\omega_4^2\right)\left(s^2+\omega_s^2\right)\ldots}{a_n\left(s^2+\omega_1^2\right)\left(s^2+\omega_3^2\right)\left(s^2+\omega_5^2\right)\ldots}. \tag{2.55}$$

Now a pole at zero requires the extraction of a shunt inductor of value (see Figure 2-9(b))

$$L_0 = \left.\frac{1}{sY\,(s)}\right|_{s=0} \tag{2.56}$$

and a pole at infinity requires the extraction of a shunt capacitor of value (see Figure 2-9(d))

$$C_\infty = \left.\frac{Y(s)}{s}\right|_{s=\infty}. \tag{2.57}$$

A pole at a finite frequency requires the extraction of a shunt-series LC block (as shown in Figure 2-9(f)) with values

$$L_i = \left.\frac{s}{(s^2+\omega_o^2)\,Y\,(s)}\right|_{s=j\omega_o} \quad \text{and} \quad C_i = \frac{1}{\omega_o^2.L_i}. \tag{2.58}$$

Many aspects of filter synthesis can seem abstract when presented in full generality. Consequently it is common to illustrate filter synthesis concepts using examples. Following this time-honored tradition, an example is now presented.

EXAMPLE 2.2 Element Extraction for a Third-Order Lowpass Filter

A third-order maximally flat filter has the reflection coefficient

$$\Gamma_1(s) = \frac{s^3}{(s+1)\left(s^2+s+1\right)}. \tag{2.59}$$

Synthesize this filter as a doubly terminated network.

Solution:

The reflection coefficient function (Equation (2.59)) has all its poles located at infinity, so the corresponding network realization must be made of simple L or C elements and terminated in a resistor. Hence, referring to Figure 2-2 and considering a 1 Ω system,

$$Z_{\text{in},1}(s) = \frac{1+\Gamma_1(s)}{1-\Gamma_1(s)} = \frac{2s^3+2s^2+2s+1}{2s^2+2s+1}. \tag{2.60}$$

Note that the input impedance approaches infinity as the frequency goes to infinity, hence a series inductor must be extracted. The value of this inductor is

$$L_{\infty 1} = \left.\frac{Z_{\text{in},1}(s)}{s}\right|_{s=\infty} = 1\ \text{H}. \tag{2.61}$$

The filter is developed by extracting one element at a time. Following the extraction of the first element, the second-stage impedance is left. Now the impedance function is

$$Z_{\text{in},2}(s) = Z_{\text{in},1}(s) - sL_{\infty 1} = \frac{2s^3 + 2s^2 + 2s + 1}{2s^2 + 2s + 1} - sL_{\infty 1}$$
$$= \frac{2s^3 + 2s^2 + 2s + 1 - s\left(2s^2 + 2s + 1\right)}{2s^2 + 2s + 1} = \frac{s+1}{2s^2 + 2s + 1}.$$

Note that the stage impedance above, $Z_{\text{in},2}$, approaches zero as the frequency goes to infinity. There is not a single series element that would cause this. However, the stage admittance function,

$$Y_{\text{in},2}(s) = \frac{1}{Z_{\text{in},2}(s)} = \frac{2s^2 + 2s + 1}{s+1}, \tag{2.62}$$

goes to infinity as the frequency approaches infinity and so a shunt capacitor is extracted:

$$Y_{\text{in},3}(s) = Y_{\text{in},2}(s) - sC_{\infty 2} = \frac{1}{s+1}, \tag{2.63}$$

where

$$C_{\infty 2} = 2 \text{ F}. \tag{2.64}$$

So sometimes it is more convenient to consider extraction of an admittance and sometimes it is better to consider extraction of an impedance.

By examining the remaining stage impedance, it is seen that a pole exists at infinity, and so a series inductor, $L_{\infty 3}$, is extracted. The value of this inductor comes from

$$Z_{\text{in},3} = \frac{1}{Y_{\text{in},3}} = s + 1\ \Omega \tag{2.65}$$

and so the inductor value is

$$L_{\infty 3} = \left.\frac{s+1}{s}\right|_{S=\infty} = 1 \text{ H}. \tag{2.66}$$

The final step is to extract a load of value 1 as follows:

$$Z_{\text{in},4} = Z_{\text{in},3} - sL_{\infty 3} = 1\ \Omega. \tag{2.67}$$

This example synthesized a doubly terminated network. The resulting network, called a ladder circuit, is shown in Figure 2-10. The left-most 1 Ω resistor is part of the source.

This circuit has a dual form consisting of two shunt capacitors separated by a series inductor. The dual circuit derives from realizing the admittance function obtained from the reflection coefficient. Other network extraction techniques are presented in Scanlan and Levy [4, 5] and Matthaei et al. [1].

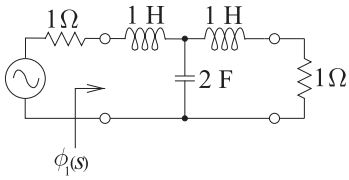


Figure 2-10: Synthesized maximally flat network with a third-order lowpass reflection response.

2.6.2 Summary

The input impedance function of a lumped-element circuit can always be expressed as the ratio of two polynomials in s and the order of the numerator and the denominator polynomials can differ by at most one [5]. If the orders differ by one, then a single inductor or capacitor can always be extracted, however, the remaining impedance function may not be realizable. This indicates that a more complex LC (and possibly R) combination is required. To be able to systematically extract arbitrarily complex circuits, a long list of possible functions, such as those shown in Figure 2-9, is required. For most of the circuits of interest the LC combinations shown in Figure 2-9 are sufficient. The next example describes impedance function extraction that requires an LC combination.

EXAMPLE 2.3 Element Extraction of an Impedance Function

Realize the impedance function $Z_w = \dfrac{4s^3 + 4s^2 + 2s + 2}{4s^2 + 2s + 1}$.

Solution:

The order of the numerator is 1 greater than the order of the denominator and this indicates that a series inductor is perhaps present. The series inductance is

$$L_1 = \left.\frac{Z_w(s)}{(s)}\right|_{z=\infty} = 1 \text{ H}. \tag{2.68}$$

The remaining impedance is

$$Z_{\text{in},2} = Z_w - sL_1 = \frac{4s^3 + 4s^2 + 2s + 2}{4s^2 + 2s + 1} - s = \frac{2s^2 + s + 2}{4s^2 + 2s + 1}. \tag{2.69}$$

The numerator and denominator of $Z_{\text{in},2}$ have the same order. Therefore a simple L or C element cannot be used to reduce the complexity of the impedance function. Thus an initial series inductor was not the right choice and the extraction must backtrack.

Figure 2-9 shows several element combinations that can be used to reduce the complexity of an impedance function. Insight into which alternative to choose comes from factoring z_w, and note that real roots are required, thus

$$Z_w = \frac{4s^3 + 4s^2 + 2s + 2}{4s^2 + 2s + 1} = \frac{(2s^2 + 1)(2s + 2)}{4s^2 + 2s + 1}. \tag{2.70}$$

Examination of Figure 2-9 reveals that a ready fit to Z_w is not found. Instead consider the admittance function

$$Y_w = \frac{1}{Z_w} = \frac{4s^2 + 2s + 1}{(2s^2 + 1)(2s + 2)}. \tag{2.71}$$

So the reduction shown in Figure 2-9(f) looks like the right candidate. The general choice for the element is

$$y_x = \frac{as}{bs^2 + 1}. \tag{2.72}$$

Choosing $b = 2$ now reduces complexity (since part of the factored denominator of Y_w now occurs), so

$$Y_w = \frac{as}{2s^2 + 1} + \left(\frac{4s^2 + 2s + 1}{(2s^2 + 1)(2s + 2)} - \frac{as}{2s^2 + 1}\right) \tag{2.73}$$

$$= \frac{as}{2s^2 + 1} + \left(\frac{(4 - 2a)s^2 + (2 - 2a)s + 1}{(2s^2 + 1)(2s + 2)}\right). \tag{2.74}$$

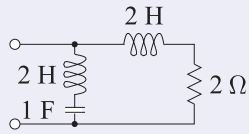
Choose $a = 1$, $$Y_w = \frac{s}{(2s^2+1)} + \frac{2s^2+1}{(2s^2+1)(2s+2)} = \frac{s}{2s^2+1} + \frac{1}{2s+2} \quad (2.75)$$

$$= \frac{s}{2s^2+1} + Y_{\text{in},2}. \quad (2.76)$$

So $C_1 L_1 = b = 1$, $C_1 = a = 1$ F, $L_1 = 2$ H, and

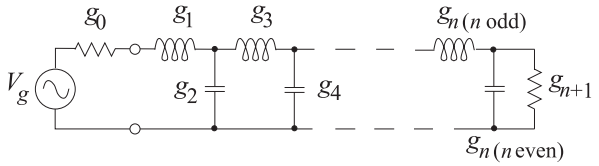
$$Y_{\text{in},2} = \frac{1}{(2s+2)} \quad \text{or} \quad Z_{\text{in},2} = \frac{1}{Y_{\text{in},2}} = 2s + 2. \quad (2.77)$$

The final network is

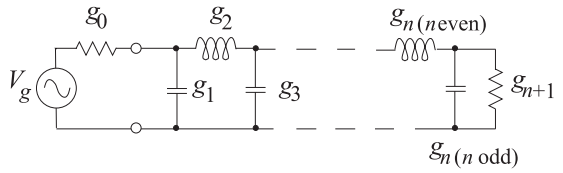


2.7 Butterworth and Chebyshev Filters

The nth-order lowpass filters constructed from the Butterworth and Chebyshev polynomials have the ladder circuit forms of Figure 2-11(a or b). Figure 2-11 uses several shorthand notations commonly used with filters. First, note that there are two prototype forms designated Type 1 and Type 2, and these are referred to as duals of each other. The two prototype forms have identical responses with the same numerical element values $g_1, \ldots, g_n$. Consider the Type 1 prototype of Figure 2-11(a). The right-most element is the resistive load, which is also known as the $(n+1)$th element. The next element to the left of this is either a shunt capacitor (of value g_n) if n is even, or a series inductor (of value g_n) if n is odd. So for the Type 1 prototype, the shunt capacitor next to the load does not exist if n is odd. The same interpretation applies to the circuit in Figure 2-11(b).



(a) Type 1



(b) Type 2

Figure 2-11: Filter prototypes in the Cauer topology. Here n is the order of the filter.

Table 2-4: Coefficients of the Butterworth lowpass prototype filter normalized to a radian corner frequency of 1 rad/s and a 1 Ω system impedance (i.e., $g_0 = 1 = g_{n+1}$).

Order, n	2	3	4	5	6	7	8	9
g_1	1.4142	1	0.7654	0.6180	0.5176	0.4450	0.3902	0.3473
g_2	1.4142	2	1.8478	1.6180	1.4142	1.2470	1.1111	1
g_3	1	1	1.8478	2	1.9318	1.8019	1.6629	1.5321
g_4		1	0.7654	1.6180	1.9318	2	1.9615	1.8794
g_5			1	0.6180	1.4142	1.8019	1.9615	2
g_6				1	0.5176	1.2470	1.6629	1.8794
g_7					1	0.4450	1.1111	1.5321
g_8						1	0.3902	1
g_9							1	0.3473
g_{10}								1

2.7.1 Butterworth Filter

A generalization of the example of the previous section leads to a formula for the element values of a ladder circuit implementing a Butterworth lowpass filter. For a maximally flat or Butterworth response the element values of the circuit in Figure 2-11(a and b) are

$$g_r = 2\sin\left\{(2r-1)\frac{\pi}{2n}\right\} \qquad r = 1, 2, 3, \ldots, n \tag{2.78}$$

and $g_0 = 1 = g_{n+1}$. Table 2-4 lists the coefficients of Butterworth lowpass prototype filters up to ninth order.

EXAMPLE 2.4 Fourth-Order Butterworth Lowpass Filter

Derive the fourth-order Butterworth lowpass prototype of Type 1.

Solution:

From Equation (2.78),

$$g_1 = 2\sin[\pi/(2\cdot 4)] = 0.765369 \text{ H} \tag{2.79}$$
$$g_2 = 2\sin[3\pi/(2\cdot 4)] = 1.847759 \text{ F} \tag{2.80}$$
$$g_3 = 2\sin[5\pi/(2\cdot 4)] = 1.847759 \text{ H} \tag{2.81}$$
$$g_4 = 2\sin[7\pi/(2\cdot 4)] = 0.765369 \text{ F}. \tag{2.82}$$

Thus the fourth-order Butterworth lowpass prototype circuit with a corner frequency of 1 rad/s is as shown in Figure 2-12.

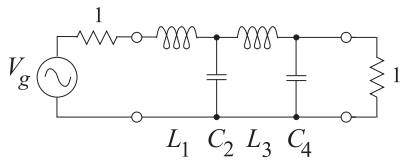


Figure 2-12: Fourth-order Butterworth lowpass filter prototype.

$L_1 = 0.765369$ H $\quad C_2 = 1.847759$ F
$L_3 = 1.847759$ H $\quad C_4 = 0.765369$ F

2.7.2 Chebyshev Filter

For a Chebyshev response, the element values of the lowpass prototype shown in Figure 2-11 are found from the recursive formula [1, 6, 7]:

$$g_0 = 1 \qquad g_1 = \frac{2a_1}{\gamma} \tag{2.83}$$

$$g_{n+1} = \left\{ \begin{array}{ll} 1, & n \text{ odd} \\ \tanh^2(\beta/4), & n \text{ even} \end{array} \right\} \tag{2.84}$$

$$g_k = \frac{4a_{k-1}a_k}{b_{k-1}g_{k-1}}, \quad k = 2,\ 3, \ldots, n \tag{2.85}$$

$$a_k = \sin\left[\frac{(2k-1)\pi}{2n}\right], \quad k = 1,\ 2, \ldots, n\,, \tag{2.86}$$

where

$$\gamma = \sinh\left(\frac{\beta}{2n}\right), \tag{2.87}$$

$$b_k = \gamma^2 + \sin^2\left(\frac{k\pi}{n}\right) \quad k = 1,\ 2, \ldots, n \tag{2.88}$$

$$\beta = \ln\left[\coth\left(\frac{R_{\text{dB}}}{2 \cdot 20\log(\text{e})}\right)\right] = \ln\left[\coth\left(\frac{R_{\text{dB}}}{17.3717793}\right)\right] \tag{2.89}$$

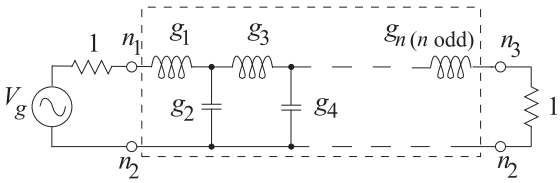
$$R_{\text{dB}} = 10\log\left(1 + \varepsilon^2\right), \tag{2.90}$$

n is the order of the filter, and ε is the ripple factor and defines the level of the ripple in absolute terms. R_{dB} is the ripple expressed in decibels (the ripple is generally specified in decibels).

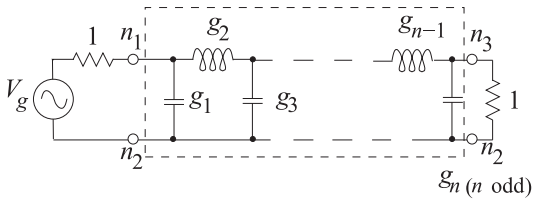
An interesting point to note here is that the source resistor, the value of which is given by g_0, and terminating resistor, the value of which is given by g_{n+1}, are only equal for odd-order filters. For an even-order Chebyshev filter the terminating resistor, g_{n+1}, will be different and a function of the filter ripple. Because it is generally desirable to have identical source and load impedances, Chebyshev filters are nearly always restricted to odd order. Thus the odd-order Chebyshev prototypes are as shown in Figure 2-13.

Also, for an odd-degree function (n is odd) there is a perfect match at DC,

$$|T(0)|^2 = 1, \tag{2.91}$$



(a) Type 1



(b) Type 2

Figure 2-13: Odd-order Chebyshev lowpass filter prototypes in the Cauer topology. Here n is the order of the filter.

Table 2-5: Coefficients of a Chebyshev lowpass prototype filter normalized to a radian corner frequency of $\omega_0 = 1$ rad/s and a 1 Ω system impedance (i.e., $g_0 = 1 = g_{n+1}$). The ripple factor, ε, is related to the ripple in decibels by Equation (2.90) (e.g., $\varepsilon = 0.1$ is a ripple of 0.0432 dB). (Note that ω_0 is the radian frequency at which the transmission response of a Chebyshev filter is down by the ripple, see Figure 2-7.)

Order	$n = 3$					
Ripple	0.01 dB	0.1 dB	0.2 dB	1.0 dB	3.0 dB	$\varepsilon = 0.1$
g_1	0.62918	1.03156	1.22754	2.02359	3.34874	0.85158
g_2	0.97028	1.14740	1.15254	0.99410	0.71170	1.10316
g_3	0.62918	1.03156	1.22754	2.02359	3.34874	0.85158

Order	$n = 5$					
Ripple	0.01 dB	0.1 dB	0.2 dB	1.0 dB	3.0 dB	$\varepsilon = 0.1$
g_1	0.75633	1.14681	1.33944	2.13488	3.48129	0.97140
g_2	1.30492	1.37121	1.33702	1.09111	0.76192	1.37208
g_3	1.57731	1.97500	2.16605	3.00092	4.53755	1.80136
g_4	1.30492	1.37121	1.33702	1.09111	0.76192	1.37208
g_5	0.75633	1.14681	1.33944	2.13488	3.48129	0.97140

Order	$n = 7$					
Ripple	0.01 dB	0.1 dB	0.2 dB	1.0 dB	3.0 dB	$\varepsilon = 0.1$
g_1	0.79694	1.18118	1.37226	2.16656	3.51852	1.00794
g_2	1.39242	1.42281	1.37820	1.11151	0.77220	1.43678
g_3	1.74813	2.09667	2.27566	3.09364	4.63898	1.93981
g_4	1.63313	1.57340	1.50016	1.17352	0.80381	1.62196
g_5	1.74813	2.09667	2.27566	3.09364	4.63898	1.93981
g_6	1.39242	1.42281	1.37820	1.11151	0.77220	1.43678
g_7	0.79694	1.18118	1.37226	2.16656	3.51852	1.00794

Order	$n = 9$					
Ripple	0.01 dB	0.1 dB	0.2 dB	1.0 dB	3.0 dB	$\varepsilon = 0.1$
g_1	0.81446	1.19567	1.38603	2.17972	3.53394	1.02347
g_2	1.42706	1.44260	1.39389	1.11918	0.76604	1.46186
g_3	1.80436	2.13455	2.30932	3.12143	4.66906	1.98372
g_4	1.71254	1.61672	1.53405	1.18967	0.81181	1.67776
g_5	1.90579	2.20537	2.37280	3.17463	4.72701	2.06485
g_6	1.71254	1.61672	1.53405	1.18967	0.81181	1.67776
g_7	1.80436	2.13455	2.30932	3.12143	4.66906	1.98372
g_8	1.42706	1.44260	1.39389	1.11918	0.76604	1.46186
g_9	0.81446	1.19567	1.38603	2.17972	3.53394	1.02347

while for an even-degree function (i.e., n is even) a mismatch exists of value

$$|T(0)|^2 = \frac{4R_L}{(R_L+1)^2} = \frac{1}{1+\varepsilon^2} \tag{2.92}$$

so that $$R_L = g_{n+1} = \left[\varepsilon + \sqrt{(1+\varepsilon^2)}\right]^2. \tag{2.93}$$

Coefficients of several Chebyshev lowpass prototype filters with different levels of ripple and odd orders up to ninth order are given in Table 2-5.

2.7.3 Summary

A Butterworth filter has a monotonic response without ripple, but a relatively slow transition from the passband to the stopband. A Chebyshev filter has a rapid transition but has ripple in either the stopband or passband. Butterworth and Chebyshev filters are special cases of **elliptical filters**, which are also called **Cauer filters**. In general, an elliptical filter has ripple in both the stopband and the passband. The level of the ripple can be selected

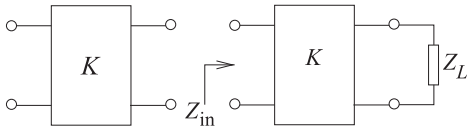


Figure 2-14: Impedance inverter (of impedance K in ohms): (a) represented as a two-port; and (b) the two-port terminated in a load.

independently in each band. With zero ripple in the stopband, but ripple in the passband, an elliptical filter becomes a **Type I Chebyshev filter**. With zero ripple in the passband, but ripple in the stopband, an elliptical filter becomes a **Type II Chebyshev filter**. With no ripple in either band the elliptical filter becomes a Butterworth filter. With ripple in both the passband and stopband, the transition between the passband and stopband can be made more abrupt or alternatively the tolerance to component variations increased.

Another type of filter is the **Bessel filter** which has maximally flat group delay in the passband, which means that the phase response has maximum linearity across the passband. The **Legendre filter** (also known as the **optimum "L" filter**) has a high transition rate from passband to stopband for a given filter order, and also has a monotonic frequency response (i.e., without ripple). It is a compromise between the Butterworth filter, with monotonic frequency response but slower transition and the Chebyshev filter, which has a faster transition but ripples in the frequency response.

More in-depth discussions of a large class of filters along with coefficient tables and coefficient formulas are available in Matthaei et al. [1], Hunter [3], Daniels [8], Lutovac et al. [9], and in most other books dedicated solely to microwave filters.

2.8 Impedance and Admittance Inverters

Inverters are two-port networks used in many RF and microwave filters. The input impedance of an inverter terminated in an impedance Z_L is $1/Z_L$. Impedance and admittance inverters are the same network, with the distinction being whether siemens or ohms are used to define them. An inverter is sometimes called a **unit element** (**UE**). At frequencies of a few hundred megahertz and below an inverter can be realized using operational and transconductance amplifiers. At microwave frequencies the simplest inverter is a one-quarter wavelength long line. In RF and microwave filter design they are used to convert a series element into a shunt element. It is much easier to realize shunt elements in distributed circuits than series elements. Similar circuit transformations enable an inductor to be replaced by a capacitor.

The schematic representation of an impedance inverter is shown in Figure, 2-14(a). The constitutive property of the inverter is that the input impedance of the terminated impedance inverter in Figure 2-14(b) is

$$Z_{\text{in}} = \frac{K^2}{Z_L}. \tag{2.94}$$

So the inverter both inverts the load impedance and scales it. Similarly, if Port 1 is terminated in Z_L the input impedance at Port 2 is Z_{in} as defined above.

An impedance inverter has the value K (in ohms), and sometimes K is called the characteristic impedance of the inverter. Sometimes K is just

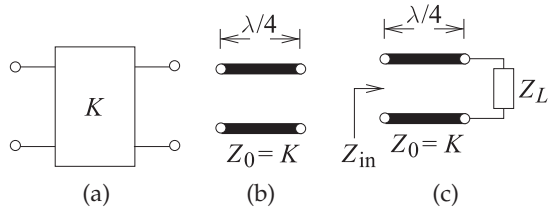


Figure 2-15: Inverter equivalence: (a) two-port impedance inverter (of impedance K): (b) a quarter-wave transmission line of characteristic impedance $Z_0 = K$; and (c) a terminated one-quarter wavelength long line.

called the impedance of the inverter. For an admittance inverter J is used and is called the characteristic admittance of the inverter, and sometimes just the admittance of the inverter. They are related as $J = 1/K$. In Section 2.4.6 of [10] it is shown that a $\lambda/4$ long line with a load has an input impedance that is the inverse of the load, normalized by the square of the characteristic impedance of the line. So an inverter can be realized at microwave frequencies using a one-quarter wavelength long transmission line (see Figure 2-15(b)). For the configuration shown in Figure 2-15(c),

$$Z_{\text{in}} = \frac{K^2}{Z_L}. \tag{2.95}$$

2.8.1 Properties of an Impedance Inverter

An impedance inverter has the $ABCD$ matrix

$$\mathbf{T} = \begin{bmatrix} 0 & \jmath K \\ \jmath/K & 0 \end{bmatrix}, \tag{2.96}$$

where K is called the characteristic impedance of the inverter. With a load impedance, Z_L (at Port 2), the input impedance (at Port 1) is (as expected)

$$Z_{\text{in}}(s) = \frac{AZ_L + B}{CZ_L + D} = \frac{\jmath K}{(\jmath/K)\, Z_L} = \frac{K^2}{Z_L}. \tag{2.97}$$

Now the $ABCD$ matrix of the transmission line of Figure 2-15(b) is

$$\begin{bmatrix} \cos\theta & \jmath Z_0 \sin\theta \\ (\jmath/Z_0)\sin\theta & \cos\theta \end{bmatrix}, \tag{2.98}$$

which is identical to Equation (2.96) when the electrical length is $\theta = \pi/2$ (i.e., when the line is $\lambda/4$ long). The inverter is shown in Figure 2-14(a) as a two-port and its implementation as a $\lambda/4$ long line is shown in Figure 2-14(c). The bandwidth over which the line realizes an impedance inverter is limited, however, as it is an ideal inverter only at the frequency at which it is $\lambda/4$ long.

2.8.2 Replacement of a Series Inductor by a Shunt Capacitor

A series inductor can be replaced by a shunt capacitor surrounded by a pair of inverters followed by a negative unity transformer (i.e., an inverter with $K = 1$). This equivalence is shown in Figure 2-16 and this will now be shown mathematically.

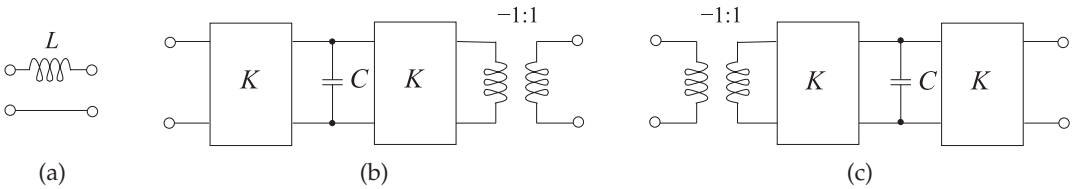


Figure 2-16: Equivalent realizations of a series inductor: (a) as a two-port; (b) its realization using a capacitor, inverters of characteristic impedance K, and a negative unity transformer; and (c) an alternative realization. $C = L/K^2$.

From Table 2-1 of [11], the $ABCD$ matrix of the series inductor shown in Figure 2-16(a) (which has an impedance of sL) is

$$\mathbf{T}_L = \begin{bmatrix} 1 & sL \\ 0 & 1 \end{bmatrix} \tag{2.99}$$

and the $ABCD$ matrix of the shunt capacitor (which has an admittance of sC) is, from Table 2-1 of of [11],

$$\mathbf{T}_1 = \begin{bmatrix} 1 & 0 \\ sC & 1 \end{bmatrix}. \tag{2.100}$$

The $ABCD$ matrix of an inverter with K in ohms (generally the unit is dropped and ohms is assumed) is

$$\mathbf{T}_2 = \begin{bmatrix} 0 & \jmath K \\ \jmath/K & 0 \end{bmatrix}, \tag{2.101}$$

and finally, the $ABCD$ matrix of a negative unity transformer, $n = -1$, is, from Table 2-1 of [11],

$$\mathbf{T}_3 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{2.102}$$

Then the $ABCD$ matrix of the cascade shown in Figure 2-16(b) is

$$\begin{aligned} \mathbf{T}_C &= \mathbf{T}_2\mathbf{T}_1\mathbf{T}_2\mathbf{T}_3 \\ &= \begin{bmatrix} 0 & \jmath K \\ \jmath/K & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ sC & 1 \end{bmatrix} \begin{bmatrix} 0 & \jmath K \\ \jmath/K & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & \jmath K \\ \jmath/K & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ sC & 1 \end{bmatrix} \begin{bmatrix} 0 & -\jmath K \\ -\jmath/K & 0 \end{bmatrix} \\ &= \begin{bmatrix} \jmath sCK & \jmath K \\ \jmath/K & 0 \end{bmatrix} \begin{bmatrix} 0 & -\jmath K \\ -\jmath/K & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & sCK^2 \\ 0 & 1 \end{bmatrix}. \end{aligned} \tag{2.103}$$

Thus $T_C = T_L$ if $L = CK^2$ (compare Equations (2.99) and (2.103)). Thus a series inductor can be replaced by a shunt capacitor with an inverter before and after it and with a negative unity transformer. The unity transformer

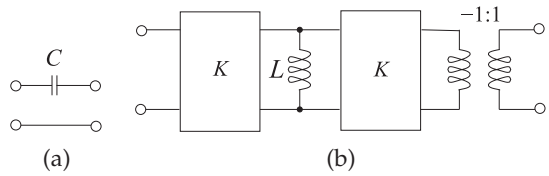


Figure 2-17: A series capacitor: (a) as a two-port; (b) its realization using a shunt inductor, inverters and negative unity transformer' $L = CK^2$.

may also be placed at the first port, as in Figure 2-16(c). Thus the two-ports shown in Figure 2-16 are all electrically identical, with the limitation being the frequency range over which the inverter can be realized. An interesting and important observation is that as a result of the characteristic impedance of the inverter (e.g., 50 Ω), a small shunt capacitor can be used to realize a large series inductance value.

EXAMPLE 2.5 Inductor Synthesis Using an Inverter

Consider the network of Figure 2-16(c) with inverters having a characteristic impedance of 50 Ω. What value of inductance is realized using a 10 pF capacitor?

Solution:

$K = 50$, so $L = CK^2 = 10^{-11} \cdot 2500 = 25$ nH.

2.8.3 Replacement of a Series Capacitor by a Shunt Inductor

A series capacitor can be replaced by a shunt inductor plus inverters and a negative transformer (see Figure 2-17). The $ABCD$ parameters of the series capacitor in Figure 2-17(a) are

$$\mathbf{T} = \begin{bmatrix} 1 & 1/sC \\ 0 & 1 \end{bmatrix}, \tag{2.104}$$

and here it is shown that the cascade in Figure 2-17(b) has the same $ABCD$ parameters. The cascade in Figure 2-17(b) has the $ABCD$ parameters

$$\begin{aligned} \mathbf{T} &= \begin{bmatrix} 0 & jK \\ j/K & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1/sL & 1 \end{bmatrix} \begin{bmatrix} 0 & jK \\ j/K & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & jK \\ j/K & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1/sL & 1 \end{bmatrix} \begin{bmatrix} 0 & -jK \\ -j/K & 0 \end{bmatrix} \\ &= \begin{bmatrix} jK/sL & jK \\ j/K & 0 \end{bmatrix} \begin{bmatrix} 0 & -jK \\ -j/K & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & K^2/sL \\ 0 & 1 \end{bmatrix}. \end{aligned} \tag{2.105}$$

So the series capacitor, C, can be realized using a shunt inductor, L, inverters, and a negative unity transformer, and $C = L/K^2$. It is unlikely that this transformation would be exploited, as much better lower-loss capacitors can be realized at RF than inductors.

2.8.4 Ladder Prototype with Impedance Inverters

The transformations discussed in the previous two sections can be used to advantage in simplifying filters. In this section the transformations are

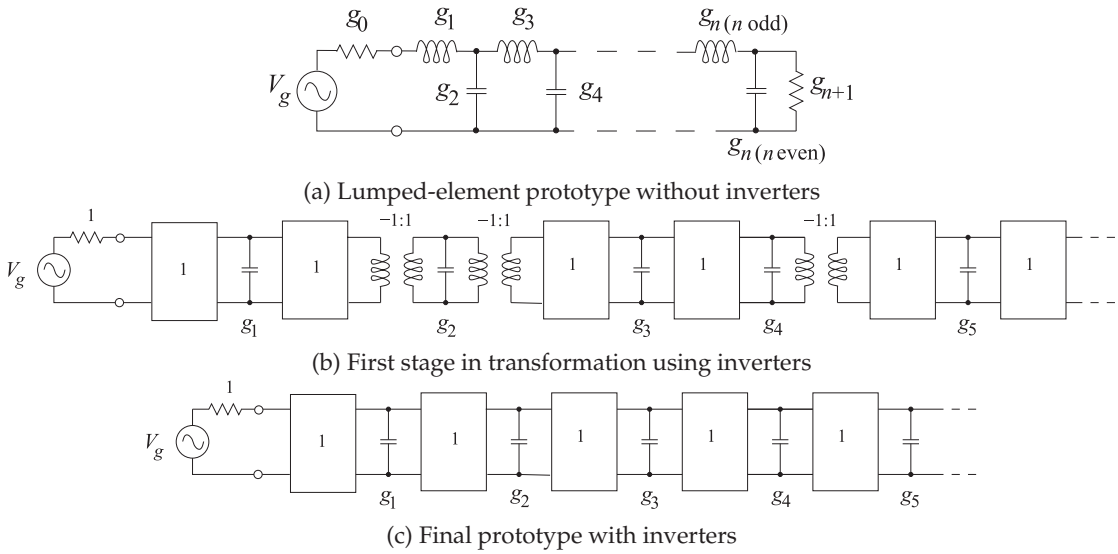


Figure 2-18: Ladder prototype filters using impedance inverters: (a) lumped-element prototype; (b) first stage in transformation using inverters; and (c) final stage.

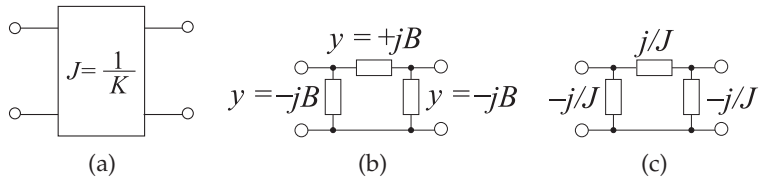


Figure 2-19: Admittance inverter: (a) as a two-port; (b) realized using lumped elements with $B = -J$; and (c) lumped equivalent circuit (the element values in (c) are impedances).

applied to the prototype lowpass ladder filter shown in Figure 2-18(a). The inductors in the ladder circuit are a particular problem as they have considerable resistance at microwave frequencies. The series inductors can be replaced by a circuit with capacitors, inverters, and transformers, as shown in Figure 2-18(b). This simplifies further to the realization shown in Figure 2-18(c), as the negative unity transformers only affect the phase of the transmission coefficient. So a lowpass ladder filter can be realized using just capacitors and inverters.

2.8.5 Lumped-Element Realization of an Inverter

The admittance inverter is functionally the same as the impedance inverter (see Figure 2-14(a)) and the schematic is the same (see Figure 2-19(a)). As will be shown, an inverter can be realized using frequency-invariant lossless elements (i.e., elements whose reactance or susceptance do not vary with frequency) using the network of Figure 2-19(b). Recall that J is used to identify an admittance inverter and K identifies an impedance inverter. If not specified by the context, the inverter (with value specified by a number) defaults to being an impedance inverter. Alternatively units can be used to indicate which type of inverter is being used. The function of

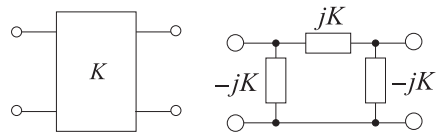


Figure 2-20: Impedance inverter: (a) as a two-port; and (b) its lumped equivalent circuit (the element values in (b) are impedances).

the inverter is the same in any case; both can be realized by one-quarter wavelength long lines, for example. For the remainder of this chapter it will be more convenient, most of the time, to use the admittance inverter, as many calculations will be in terms of admittances since most lumped elements in filters synthesis will be in shunt.

Now it will be shown that the lumped-element network of Figure 2-19(b) realizes an inverter. To do this the inverter and the lumped-element network must have the same two-port parameters. First, the $ABCD$ matrix of an inverter of characteristic admittance J is

$$\mathbf{T}_J = \begin{bmatrix} 0 & \jmath/J \\ \jmath J & 0 \end{bmatrix}. \tag{2.106}$$

Referring to Table 2-1 of [11], the circuit of Figure 2-19(b) has the $ABCD$ matrix

$$\begin{aligned} T &= \begin{bmatrix} 1 & 0 \\ -\jmath B & 1 \end{bmatrix} \begin{bmatrix} 1 & 1/(\jmath B) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\jmath B & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1/(\jmath B) \\ -\jmath B & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\jmath B & 1 \end{bmatrix} = \begin{bmatrix} 0 & -\jmath/B \\ -\jmath B & 0 \end{bmatrix}, \end{aligned} \tag{2.107}$$

where B is the susceptance of the frequency-invariant elements. Equation (2.107) is identical to Equation (2.106) if $B = -J$. More practical equivalents of the circuit of Figure 2-19(b) can be derived, as shown later.

For completeness, the lumped-element equivalent of the impedance inverter is shown in Figure 2-20 (derived from Figure 2-19 with $J = 1/K$).

EXAMPLE 2.6 Lumped Inverter Analysis

Demonstrate that Figure 2-19(b) is a lumped-element admittance inverter.

Solution:

Terminating the network in Figure 2-19(b) results in the network shown in Figure 2-21(a). This is relabeled in Figure 2-21(b), where the elements are admittances. Then

$$y_{\text{in}} = y_3 // (y_1 \$ (y_2 + y_L)), \tag{2.108}$$

where // indicates "in parallel with" and \$ indicates "in series with." These are common shorthand notations in circuit calculations. Continuing on from Equation (2.108),

$$y_{\text{in}} = y_3 + \left[\frac{1}{y_1} + \left(\frac{1}{y_2 + y_L}\right)\right]^{-1} = y_3 + \left[\frac{y_2 + y_L + y_1}{y_1(y_2 + y_L)}\right]^{-1}. \tag{2.109}$$

Substituting $y_1 = \jmath B$ and $y_2 = y_3 = -\jmath B$, this becomes

$$y_{\text{in}} = -\jmath B + \frac{\jmath B(y_L - \jmath B)}{y_L} = \frac{-\jmath B y_L + \jmath B y_L + B^2}{y_L} = \frac{B^2}{y_L}. \tag{2.110}$$

Thus the lumped-element circuit of Figure 2-19(b) is a lumped-element admittance inverter of value B (in siemens).

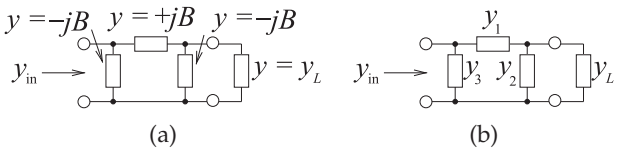


Figure 2-21: Terminated lumped-element admittance inverter.

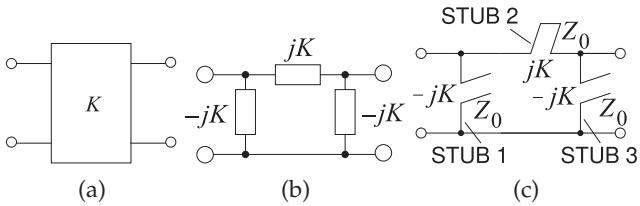


Figure 2-22: Narrowband inverter equivalents at frequency f_0: (a) impedance inverter with characteristic impedance K; (b) lumped-element equivalent network; and (c) inverter realized by short- and open-circuited stubs.

2.8.6 Narrowband Realization of an Inverter Using Transmission Line Stubs

In this section it will be shown that an impedance inverter can be implemented using short- and open-circuited stubs. The match is good over a narrow band centered at frequency f_0. An impedance inverter is shown in Figure 2-22(a) and its equivalent lumped-element network is shown in Figure 2-22(b). A stub-based implementation is shown in Figure 2-22(c), where there are short- and open-circuited stubs of characteristic impedance Z_0. The input impedance of the stubs is shown at the inputs of the stubs. The stubs have an electrical length θ at f_0 and the stubs are one-quarter wavelength long (i.e., resonant) at what is called the commensurate frequency, f_r.

Now it will be shown that the network of Figure 2-22(c) is a good representation of the inverter at f_0. This is done by matching $ABCD$ parameters. The $ABCD$ parameter matrix of an inverter is

$$T = \begin{bmatrix} 0 & \jmath K \\ \jmath/K & 0 \end{bmatrix} \tag{2.111}$$

and, at frequency f_0, the $ABCD$ parameter matrix of the stub circuit of Figure 2-22(c) is

$$\begin{aligned} T &= \begin{bmatrix} 1 & 0 \\ -1/[\jmath Z_0 \tan(\theta)] & 1 \end{bmatrix} \begin{bmatrix} 1 & \jmath Z_0 \tan(\theta) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/[\jmath Z_0 \tan(\theta)] & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & \jmath Z_0 \tan(\theta) \\ -1/[\jmath Z_0 \tan(\theta)] & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/[\jmath Z_0 \tan(\theta)] & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & \jmath Z_0 \tan(\theta) \\ \jmath/[Z_0 \tan(\theta)] & 0 \end{bmatrix}. \end{aligned} \tag{2.112}$$

Thus, equating Equations (2.111) and (2.112), the stub network is a good representation of the inverter if

$$K = Z_0 \tan(\theta), \tag{2.113}$$

and so the required characteristic impedance of each stub at frequency f_0 is

$$Z_0 = \frac{K}{\tan(\theta)} = \frac{K}{\tan\left(\frac{\pi}{2}\frac{f_0}{f_r}\right)}. \tag{2.114}$$

Special Case, $f_r = 2f_0$

In most designs the stub resonant frequency, f_r (also called the commensurate frequency), is chosen to be twice that of the center frequency of the design, f_0. So with $f_r = 2f_0$, then at f_0

$$Z_0 = \frac{K}{\tan\left(\frac{\pi}{2}\frac{f_0}{2f_0}\right)} = \frac{K}{\tan \pi/4} = K \tag{2.115}$$

and the input impedance of the stub is $\jmath K$. So the characteristic impedance of the transmission line stub is $Z_0 = K$.

2.9 Filter Transformations

So far the discussion has centered around lowpass filters. Filter design technology has developed so that the design of a corresponding lowpass filter is the essential first step, and this is used as a prototype to derive a filter with other characteristics. The three most important transformations are listed here:

1. Impedance scaling: A lowpass filter prototype is referenced to a standard impedance. Usually $1\ \Omega$ is used, so the reference source and load resistances are also $1\ \Omega$. To reference to a higher or lower impedance, scaling of the impedance of all elements of the filter is required.
2. Corner frequency scaling: The corner frequency of a lowpass filter prototype is normalized to 1 rad/s. To reference to another frequency, the values of the elements must be altered so that they have the same impedance at the scaled frequency.
3. Filter type transformation: These transformations enable the circuit of a lowpass filter to be converted to a circuit with a bandpass, bandstop, or highpass response. The concept is that the response at DC is to be replicated at infinite frequency for the highpass (so capacitors become inductors, etc.); to be replicated at the center of the passband for a bandpass filter (so capacitors become a shunt LC circuit); and to be inverted at the center of a bandstop filter (so capacitors become a series LC circuit).

The transformations can be performed in any order. The key point is that the complexity of the lowpass circuit design is minimal since the number of elements does not grow until the filter type transformation is performed, so this step is often done last.

2.9.1 Impedance Transformation

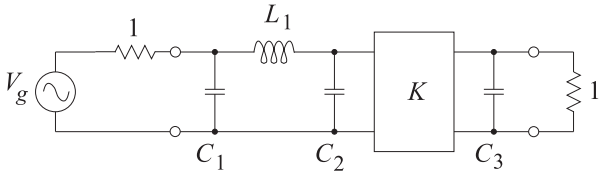
The lowpass prototypes discussed up to now have been referred to a $1\ \Omega$ system. The system impedance can be changed to any level by simply scaling the impedances of all the circuit elements in a filter by the same amount, as shown in Figure 2-23. The impedance transformation follows the same procedure for all filter types (the procedure is the same for lowpass, highpass, bandpass, and bandstop filters, for example). It should also be noted that impedance scaling of a transmission line is simply multiplication of the line's characteristic impedance by the scale factor.

(a) Impedance transformation of an inductor

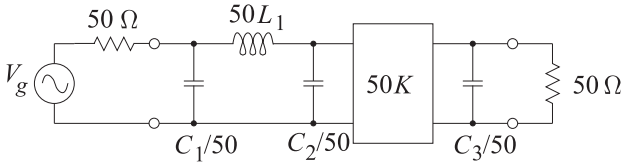
(b) Impedance transformation of a capacitor

(c) Impedance transformation of an inverter

Figure 2-23: Impedance transformations. The impedances of the elements are increased by a factor Z_0.



(a) Filter normalized to a reference impedance of 1 Ω



(b) Filter impedance transformed to a reference impedance of 50 Ω

Figure 2-24: Impedance transformation of an example filter.

EXAMPLE 2.7 Lowpass Filter Design

Consider the lowpass filter, with an inverter, shown in Figure 2-24(a). This filter is referenced to 1 Ω, as the source and load impedances are both 1 Ω. Redesign the filter so that the same frequency response is obtained with 50 Ω source and load impedances.

Solution:

It is necessary to impedance transform from 1 Ω to 50 Ω. The resulting filter is shown in Figure 2-24(b). Each element has an impedance (resistance or reactance) that is 50 times larger than it had in the 1 Ω prototype.

2.9.2 Frequency Transformation: Lowpass

Frequency transformations differ depending on the type of the filter. So it is normal to frequency transform the lowpass prototype before converting the prototype to another form (such as bandpass). The lowpass prototypes normally have a band-edge or cutoff frequency at an angular frequency of unity, that is, 1 rad/s. The band-edge frequency can be transformed from unity to an arbitrary angular frequency ω_c, as shown in Figure 2-25, by scaling the reactive elements, as shown in Figures 2-26 and 2-27(a), so that

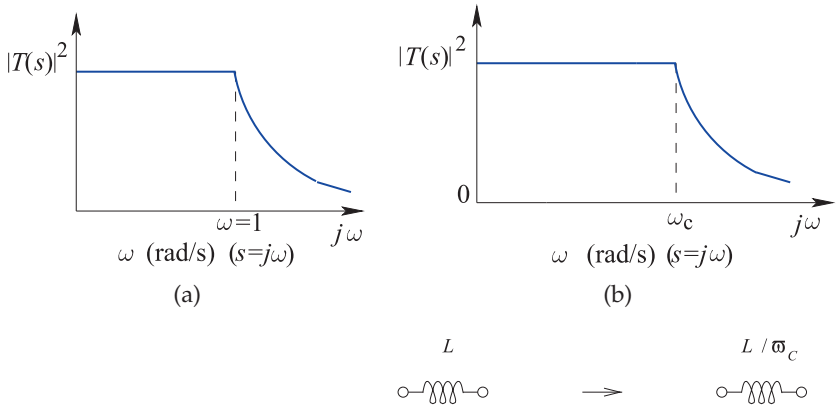


Figure 2-25: Frequency transformation of a lowpass filter response from (a) one normalized to a corner frequency of 1 rad/s, to (b) one with a radian corner frequency of ω_c.

(a) Frequency transformation of an inductor

(b) Frequency transformation of a capacitor

Figure 2-26: Frequency transformations. The impedance of the new (scaled) element at the new (scaled) frequency is the same as it was at the original frequency.

they have the same impedance at the transformed frequency as they did at the original frequency. The inverter is unchanged, as are the source and load resistances, since these are frequency independent.

2.9.3 Lowpass to Highpass Transformation

The transformation of a lowpass filter prototype to a highpass filter is shown diagrammatically in Figure 2-28. Mathematically ω in the transfer function of the lowpass prototype is replaced by $-1/\omega$, that is

$$T_{\text{highpass}}(\omega) = T_{\text{lowpass}}(-1/\omega). \tag{2.116}$$

In terms of a lumped element in the lowpass prototype circuit, if an element has an impedance $j\omega L$ and L is frequency independent, then the corresponding element in the highpass prototype filter has an impedance $1/(j\omega^2 L)$. Thus reactive elements are transformed as shown in Figure 2-27(b), where ω_0 is the corner frequency of both the lowpass and highpass prototype circuits. So inductors are transformed into capacitors and capacitors into inductors. For example, the odd-order lowpass filter prototypes shown in Figure 2-13 are transformed into the highpass filters shown in Figure 2-29.

2.9.4 Lowpass to Bandpass Transformation

Understanding the transformation of the lowpass filter into its corresponding bandpass form requires that the lowpass filter be considered with both its positive and negative frequency responses, as shown in Figure 2-30(a). This response is shifted in frequency to obtain the bandpass response shown in Figure 2-30(b). Mathematically the radian frequency, ω, in the response

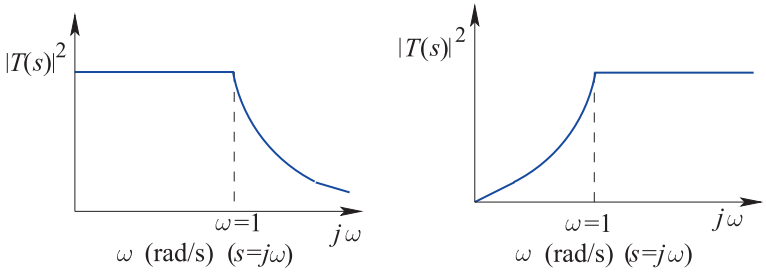
(a) LOWPASS PROTOTYPE TO LOWPASS TRANSFORMATION		
Lowpass prototype	Lowpass element	Reactance transformation ω_0 = corner frequency
C_0	C_1	$C_1 = C_0/\omega_0$
L_0	L_1	$L_1 = L_0/\omega_0$

(b) LOWPASS PROTOTYPE TO HIGHPASS TRANSFORMATION		
Lowpass prototype	Highpass element	Reactance transformation ω_0 = corner frequency
C_0	L_1	$L_1 = 1/(\omega_0 C_0)$
L_0	C_1	$C_1 = 1/(\omega_0 L_0)$

(c) LOWPASS PROTOTYPE TO BANDPASS TRANSFORMATION		
Lowpass prototype	Bandpass element	Reactance transformation ω_0 = center of passband
C_0	L_1 in parallel with C_1	$C_1 = \alpha C_0/\omega_0 \quad L_1 = 1/(\alpha C_0 \omega_0)$
L_0	L_1 in series with C_1	$L_1 = \alpha L_0/\omega_0 \quad C_1 = 1/(\alpha L_0 \omega_0)$

(d) LOWPASS PROTOTYPE TO BANDSTOP TRANSFORMATION		
Lowpass prototype	Bandstop element	Reactance transformation ω_0 = center of stopband
C_0	L_1 in series with C_1	$L_1 = \alpha/(C_0\omega_0) \quad C_1 = C_0/(\omega_0\alpha)$
L_0	L_1 in parallel with C_1	$L_1 = L_0/(\alpha\omega_0) \quad C_1 = \alpha/(L_0\omega_0)$

Figure 2-27: Transformations of the elements of a prototype lowpass filter to obtain specific filter types. The corner frequency of the lowpass prototype is 1 rad/s. In the transformations to bandpass and bandstop filters, $\omega_0 = 1/\sqrt{L_1 C_1} = \sqrt{\omega_1 \omega_2}$; ω_1 and ω_2 are the band-edge frequencies, and α is the transformation constant, $\alpha = \omega_0/(\omega_2 - \omega_1)$.



(a) Lowpass response (b) Highpass response

Figure 2-28: Lowpass to highpass transformation.

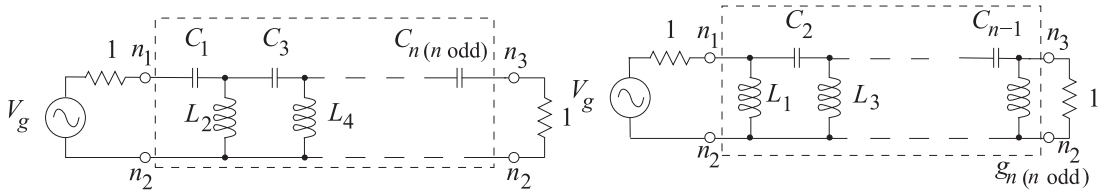


Figure 2-29: Odd-order Chebyshev highpass filter prototypes in the Cauer topology. Here n is the order of the filter.

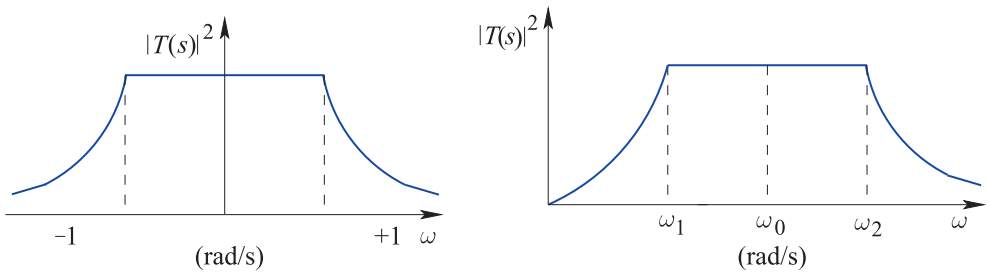


Figure 2-30: Frequency responses in lowpass to bandpass transformation ($s = \jmath\omega$).

function is replaced by its bandpass form,

$$\omega \to \left[\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right]. \tag{2.117}$$

That is, $$T_{\text{bandpass}}(\omega) = T_{\text{lowpass}}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right), \tag{2.118}$$

and so the transfer function of the bandpass filter is derived from the transfer function of the bandpass filter with ω replaced by $(\omega/\omega_0 - \omega_0/\omega)$. This separately maps the -1 and $+1$ band-edge radian frequencies of the lowpass response to the bandpass frequencies ω_1 and ω_2:

$$-1 \to \left[\frac{\omega_1}{\omega_o} - \frac{\omega_o}{\omega_1}\right] \quad \text{and} \quad +1 \to \left[\frac{\omega_2}{\omega_o} - \frac{\omega_o}{\omega_2}\right]. \tag{2.119}$$

Solving the above equations simultaneously yields the center frequency ω_0 and the band-edge frequencies ω_1 and ω_2 with

$$\omega_o = \sqrt{\omega_1\omega_2} \tag{2.120}$$

and the so-called transformation constant

$$\alpha = \frac{\omega_o}{\omega_2 - \omega_1}. \tag{2.121}$$

The resulting element conversions are given in Figure 2-27(c).

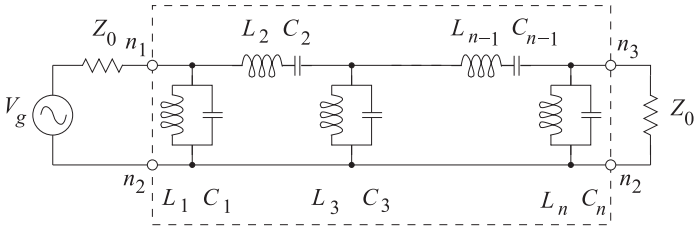
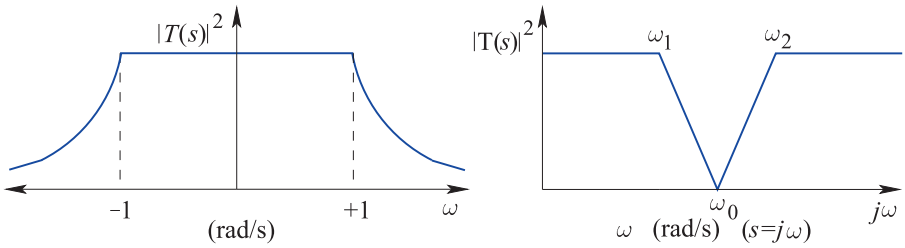


Figure 2-31: Lumped-element odd-order (nth-order) Chebyshev bandpass filter prototypes in the Type II Cauer topology.



(a) Two-sided lowpass filter response (b) Bandstop filter approximation

Figure 2-32: Frequency responses in lowpass to bandstop transformation.

As an example, a lumped-element Type 2 Cauer bandpass filter is shown in Figure 2-31. The shunt LC combination and the series LC combinations are resonators resonant at the center frequency of the filter. Here the filter is normalized to Z_0 source and load impedances. As these are the same, this filter topology only applies for an odd-order filter. Combining transformations, the element values of a lumped bandpass filter with center radian frequency $\omega_0 = 2\pi f_0$ and radian bandwidth $\omega_{\text{BW}} = 2\pi(f_2 - f_1)$ are as follows (g_r is from the lowpass prototype):

$$C_r = \begin{cases} \dfrac{g_r}{\omega_{\text{BW}} Z_0} & r = \text{odd} \\ \dfrac{\omega_{\text{BW}}}{\omega_0^2 g_r Z_0} & r = \text{even} \end{cases} \quad \text{and} \quad L_r = \begin{cases} \dfrac{\omega_{\text{BW}} Z_0}{\omega_0^2 g_r} & r = \text{odd} \\ \dfrac{g_r Z_0}{\omega_{\text{BW}}} & r = \text{even} \end{cases} . \tag{2.122}$$

2.9.5 Lowpass to Bandstop Transformation

Again, consider both the positive and negative frequency responses of the lowpass filter prototype, as shown in Figure 2-32(a). This response is shifted in frequency to obtain the bandstop response shown in Figure 2-32(b). Mathematically the frequency, ω, in the response function is replaced by its bandstop form:

$$\omega \to \left[\alpha \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right) \right]^{-1} . \tag{2.123}$$

That is, $$T_{\text{bandstop}}(\omega) = T_{\text{lowpass}} \left(\frac{1}{\alpha} \left[\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right]^{-1} \right) . \tag{2.124}$$

The center frequency (corresponding to DC in the lowpass prototype response) is

$$\omega_o = \sqrt{\omega_1 \omega_2} \tag{2.125}$$

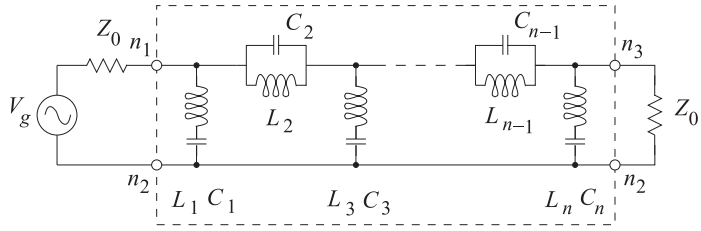


Figure 2-33: Lumped-element odd-order (nth-order) Chebyshev bandstop filter prototypes in the type II Cauer topology.

and the transformation constant is

$$\alpha = \frac{\omega_o}{\omega_2 - \omega_1}, \tag{2.126}$$

where ω_1 and ω_2 are the band-edge radian frequencies. The resulting element conversions are given in Figure 2-27(d).

Combining transformations, the element values of a lumped bandstop filter with center radian frequency $\omega_0 = 2\pi f_0$ and radian bandwidth $\omega_{\rm BW} = 2\pi(f_2 - f_1)$ are as follows:

$$C_r = \begin{cases} \dfrac{g_r \omega_{\rm BW}}{\omega_0^2 Z_0} & r = \text{odd} \\ \dfrac{1}{\omega_{\rm BW} g_r Z_0} & r = \text{even} \end{cases} \quad \text{and} \quad L_r = \begin{cases} \dfrac{Z_0}{\omega_{\rm BW} g_r} & r = \text{odd} \\ \dfrac{g_r \omega_{\rm BW} Z_0}{\omega_0^2} & r = \text{even} \end{cases}. \tag{2.127}$$

A lumped-element type II Cauer bandstop filter is shown in Figure 2-33. The parallel LC combination and the series LC combinations are resonators resonant at the center frequency of the filter. The parallel LC resonator is an open circuit at the center frequency of the stopband and the series LC resonators are short circuits. The LC resonators are implemented using resonators, usually transmission line segments and not lumped components.

2.9.6 Transformed Ladder Prototypes

Combining the filter type transformations, and with appropriate use of inverters, the original lowpass prototype ladder filter and its various filter type transforms are shown in Figure 2-34.

2.10 Cascaded Line Realization of Filters

In this section, filters are presented that use cascaded sections of transmission line, realizing series inductors and shunt capacitors. For example, a short (less than one-quarter wavelength long) length of relatively high-impedance line behaves predominantly as a series inductance. Also, a very short (much less than one-quarter wavelength long) length of relatively low-impedance line acts predominantly as a shunt capacitance. So a Pi network of lumped elements can be realized with alternate sections of low- and high-impedance microstrip lines. Such an inductive line is shown in Figure 2-35(a), which has the two equivalent circuits shown in Figure 2-35(b and c). Basic transmission line theory gives the input reactance of the line of length, ℓ (with a low-impedance load),

$$X_L = Z_0 \sin(2\pi\ell/\lambda_g), \tag{2.128}$$

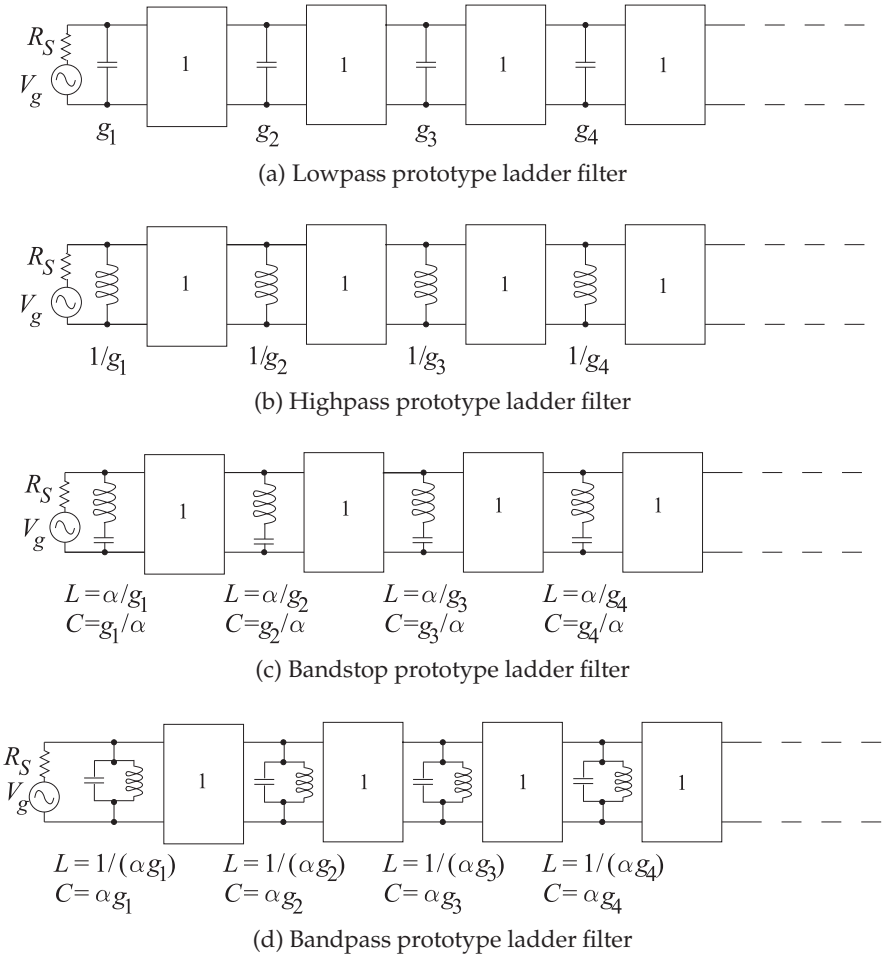


Figure 2-34: Ladder prototype filters.

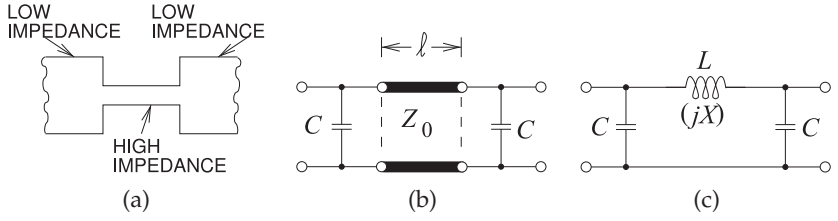


Figure 2-35: Inductive length of line with adjacent capacitive lines: (a) microstrip form; (b) lumped equivalent circuit; and (c) lumped-distributed equivalent circuit.

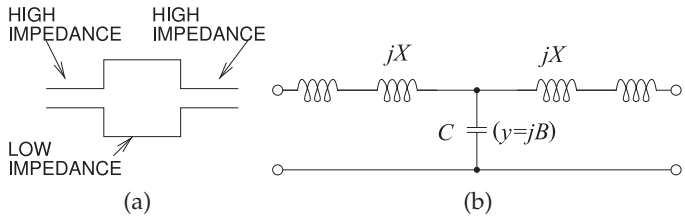
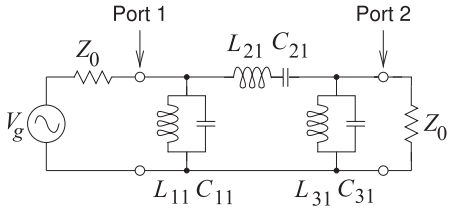


Figure 2-36: Capacitive length of line with adjacent inductive lines: (a) microstrip; and (b) lumped equivalent (the left-most and right-most series inductors come from the high-impedance lines).



Butterworth	Chebyshev
$C_{11} = C_{31} = 31.8310$ pF	$C_{11} = C_{31} = 41.5369$ pF
$L_{11} = L_{31} = 795.775$ pH	$L_{11} = L_{31} = 609.824$ pH
$C_{21} = 159.155$ fF	$C_{21} = 188.299$ fF
$L_{21} = 159.155$ nH	$L_{21} = 134.522$ nH
	ripple, $\varepsilon = 0.1 = 0.043$ dB

Figure 2-37: Lumped-element 3rd-order bandpass filters in a $Z_0 = 50\ \Omega$ system with center frequency $f_0 = 1$ GHz and a 3 dB bandwidth of 10%.

so that the length of this predominantly inductive line is

$$\ell = \frac{\lambda_g}{2\pi} \sin^{-1}\left(\frac{\omega L}{Z_0}\right). \tag{2.129}$$

Previously it was shown that a short length of line having a relatively low characteristic impedance yields a capacitive element and this is shown, together with its equivalent circuit, in Figure 2-36. The predominating shunt capacitance is determined by first considering the susceptance

$$B = \frac{1}{Z_0} \sin\left(\frac{2\pi\ell}{\lambda_g}\right) \tag{2.130}$$ so that $$\ell = \frac{\lambda_g}{2\pi} \sin^{-1}(\omega C Z_0). \tag{2.131}$$

Thus a cascade of low-impedance and high-impedance lines can realize (approximately) an LC ladder network.

It is tempting to take filters synthesized in lumped-element form and realize them in the above manner with transmission lines. The series element realization concept presented in this section could be extended using shorted and open stubs to better realize shunt elements. The problem with this approach is that the resulting filters are narrowband and the response outside the desired operating range is unpredictable. It is far better to employ the Richards's transformation, considered in Section 2.12, which is a much broader bandwidth technique for realizing filters in distributed form.

2.11 Butterworth and Chebyshev Bandpass Filters

Designs of Butterworth and Chebyshev filters with center frequencies of 1 GHz and 3 dB bandwidths of 10% are shown in Figure 2-37. Their transmission, S_{21}, and reflection, S_{11}, responses are shown in Figure 2-38. The filter skirts, i.e. the transitions from the passband to the stopbands, is steeper for the Chebyshev filter than for the Butterworth filter as expected. The ripple of the Chebyshev filter is very small and not seen in this plot. In a real filter there will be loss and low-level ripples in the magnitude response

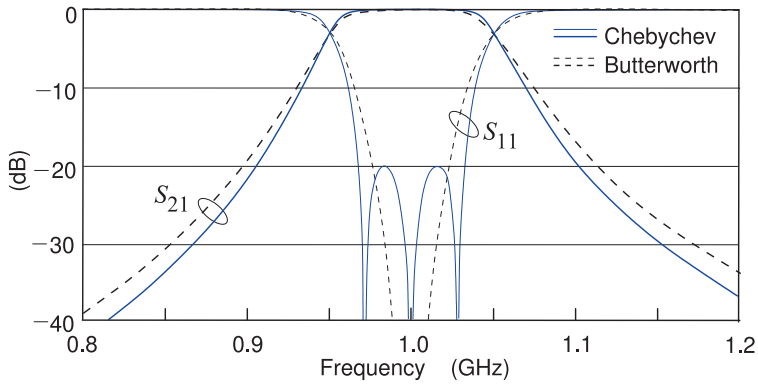


Figure 2-38: Insertion loss (S_{21}) and return loss (S_{11}) of the Butterworth and Chebychev lumped-element bandpass filters in a 50 Ω system.

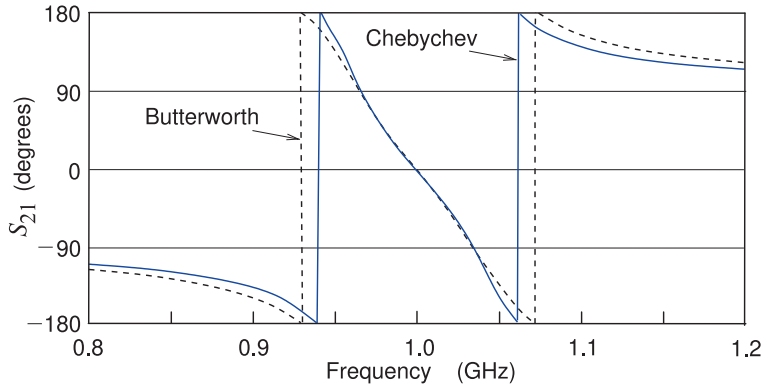
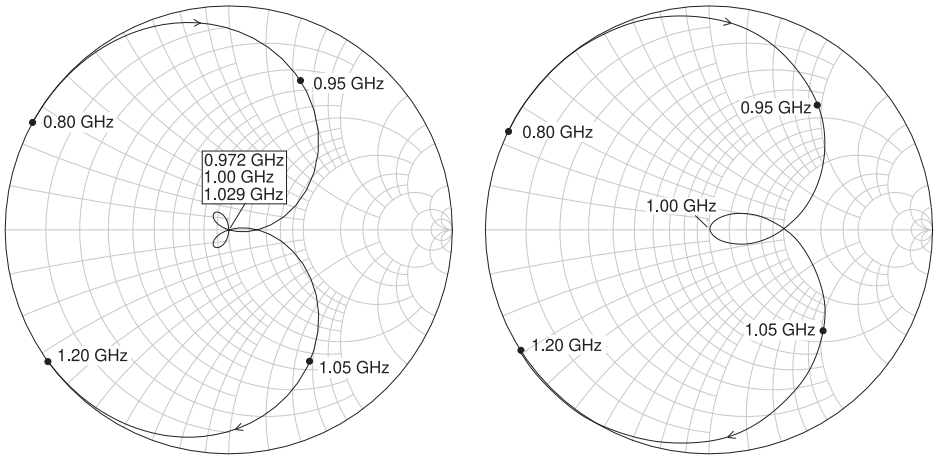


Figure 2-39: Phase of the transmission response (S_{21}) of the Butterworth and Chebyshev lumped-element filters. The discontinuities seen in the phase from $-180°$ to $+180°$ are artifacts and the phases of the filters reduce monotonically with increasing frequency. (For example, the phase at 0.8 GHz is $-110° + 360° = 250°$.)

disappear. However even with loss the steep skirts of the Chebyshev transmission response remain. Also, even with loss, the impact of the Chebyshev ripples is clearly seen in the reflection (S_{11}) response. In Figure 2-38 three distinct S_{11} zeros are seen and these correspond to the three poles of the Chebyshev filter's S_{21} response, but of course we cannot see these. (In the Laplace transfer function there are 3 complex poles each pair being transformed from one of the three poles of the lowpass prototype.) The Butterworth filter also has three (complex) S_{11} zeros and these are all at the center frequency of the bandpass filter, 1 GHz.

Another characteristic that differs between the Chebyshev and Butterworth responses is seen in their phase responses plotted in Figure 2-39. Each pole in the S_{21} characteristic causes a $90°$ phase change. The three complex poles (i.e. six actual poles) of S_{21} then result in six $90°$ phase changes in S_{21} for a total phase change of $450°$. Small ripples are seen in the Chebyshev phase responses in the pass band while the phase changes for the Butterworth filter are smooth. (The Chebyshev phase ripples remain even with low-level loss.)

The magnitude and phase responses, Figures 2-38 and 2-39, do not provide complete visualization of the filter characteristics. Additional insight is provided in the S_{11} loci on the Smith chart, see Figure 2-40. Figure 2-40 shows the S_{11} characteristics plotted on Smith charts. (In the passband the S_{21} locus would be very close to the unit circle and little of value is observed in the passband.) There is a wealth of information here. First consider the response for the Chebyshev filter, Figure 2-40(a). As frequency increases from 0.8 GHz, the locus of S_{11} is first close to the unit circle and then approaches the origin



(a) Chebyshev (b) Butterworth

Figure 2-40: Smith chart plot of S_{11} of the Butterworth and Chebychev lumped-element filters.

of the polar plot as the frequency approaches the passband of the filter. A special characteristic of the Chebyshev response is the looping which here results in three passes of the locus through the origin. These are the three zeros. Eventually the locus of S_{11} increases as frequency increases above the passband. A loop is also seen in the Butterworth response, Figure 2-40(b). This loop goes through the origin and while it seems that there is just one zero there are actually three. What is happening is that as frequency increases and as the locus of the Butterworth S_{11} response approaches the origin, the movement of the locus with respect to frequency slows down. The best way to be convinced that there are three zeros of S_{11} is to look at the transmission phase ($\angle S_{21}$) response in Figure 2-38.

Comprehensive visualization of the filter response requires the rectangular plots of the magnitudes of S_{21} and S_{11}, Figure 2-38, of S_{21} phase, Figure 2-39, and the Smith chart plot of S_{11}, Figure 2-40. The phase plot convinces you of the number of zeros in your design which is important in interpreting the Butterworth results. A physical implementation of the design will not be exact and so tuning is required. Then the most important characterizations are the rectangular magnitude and Smith chart responses. Here the loops on the Smith chart, even if they do not go through the origin exactly, distinguish the Butterworth and Chebyshev responses. Matching may be required to shift the loops to the origin of the polar plot.

2.12 Richards's Transformation

Richards's transformation is a remarkable scheme that takes into account the actual properties of transmission lines, yielding broadband transmission line-based implementations of lumped-element filter prototypes [12–15].

2.12.1 Richards's Transformation and Transmission Lines

Consider a section of transmission line of electrical length θ with $ABCD$ parameters

$$T = \begin{bmatrix} \cos(\theta) & \jmath Z_0 \sin(\theta) \\ \jmath/Z_0 \sin(\theta) & \cos(\theta) \end{bmatrix}. \tag{2.132}$$

If this line is terminated in a load, Z_L, then its input impedance is

$$Z_{\text{in}}(\theta) = \frac{\cos(\theta) Z_L + \jmath Z_0 \sin(\theta)}{j/Z_0 \sin(\theta) Z_L + \cos(\theta)}. \tag{2.133}$$

Now examine two extreme conditions. As the load impedance increases, eventually becoming an open circuit, the input impedance of a line with electrical length θ is defined in terms of a cotangent of the electrical length:

$$Z_L \to \infty \quad \Rightarrow Z_{\text{in}}(\theta) = \frac{Z_0}{\jmath} \cot(\theta) \tag{2.134}$$

$$Y_{\text{in}}(\theta) = \jmath Y_0 \tan(\theta). \tag{2.135}$$

As the load impedance reduces to become a short circuit, the input impedance of a line with electrical length θ is defined in terms of a tangent of the electrical length:

$$Z_L \to 0 \quad \Rightarrow \quad Z_{\text{in}}(\theta) = \jmath Z_0 \tan(\theta). \tag{2.136}$$

These results lead to the Richards's transformation, which replaces the Laplace variable, s, by Richards's variable, S, where $S = \jmath\alpha \tan(\theta)$. This transformation is written

$$s \to S = \jmath\alpha \tan(\theta). \tag{2.137}$$

For now α and θ are constants that can be chosen as design variables. θ, of course, is the electrical length of the line. Also, α must have the units of impedance and it is the characteristic impedance of the transmission line.

Applying the Richards's transformation to a capacitor, the admittance of the element is transformed as follows:

$$y = sC \to Y = SC = \jmath\alpha C \tan(\theta), \tag{2.138}$$

so that the capacitor is transformed into an open-circuited stub with characteristic admittance

$$Y_0 = \alpha C\ . \tag{2.139}$$

If a lumped-element capacitor with admittance $y = sC$ is to be realized using a transmission line, the admittance $Y = SC = \jmath\alpha C \tan(\theta)$ is instead realized. There are two parameters to select to realize this admittance. The first, α, is the characteristic admittance of the transmission line (and for any given transmission line topology there is a minimum and maximum characteristic admittance or impedance that can be realized), and θ is the electrical length of the line.

Applying the transformation to an inductor, the impedance of the element is transformed as follows:

$$Z = sL \to Z = SL = \jmath\alpha L \tan(\theta), \tag{2.140}$$

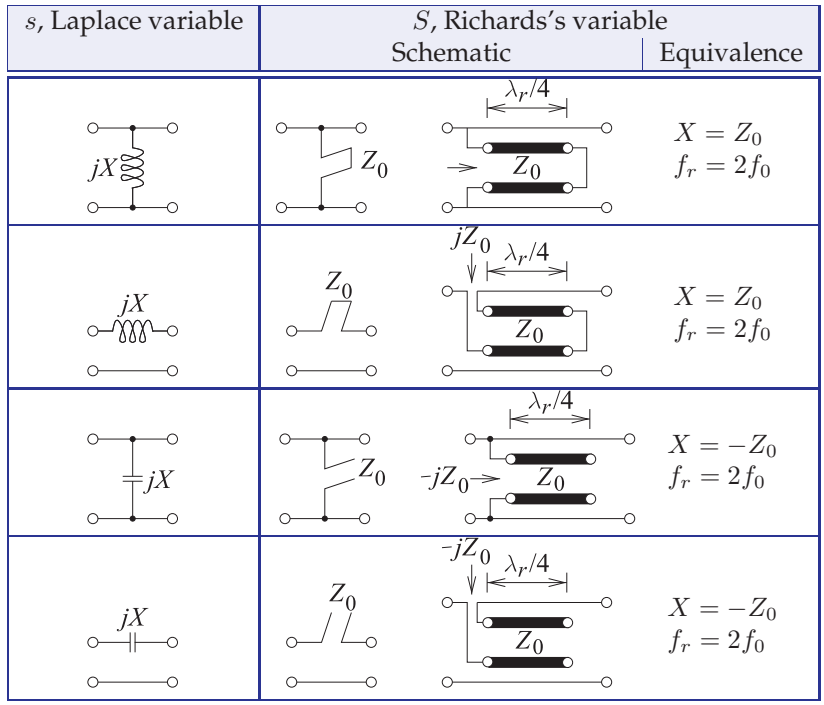


Figure 2-41: Equivalences resulting from Richards's transformation. With $f_r = 2f_0$ the transmission line stubs are one-eighth wavelength long at f_0.

so that the inductor is transformed into a short-circuited stub with characteristic impedance

$$Z_0 = \alpha L. \tag{2.141}$$

Thus the Richards's transform converts an inductor into a short-circuited stub and a capacitor into an open-circuited stub.

2.12.2 Richards's Transformation and Stubs

There is a duality between stubs and inductors and capacitors; they are coupled by Richards's transformation. One of the important quantities used in the transformation is the commensurate frequency, f_r, which most often is chosen as twice the operating frequency, f_0. Considering stubs that are one-quarter wavelength long at f_r, the duality is as shown in Figure 2-41.

2.12.3 Richards's Transformation Applied to a Lowpass Filter

In this section, Richards's transformation is used to realize a lumped-element filter in distributed form. The design example begins by considering a Chebyshev lowpass filter with the transmission characteristic

$$|T(s)|^2 = \frac{1}{1 + \varepsilon^2 |K(s)|^2}. \tag{2.142}$$

With the Richards's transformation ($s \to \jmath\omega \to \jmath\alpha \tan(\theta)$) this becomes

$$|T(\jmath\alpha \tan\theta)|^2 = \frac{1}{1 + \varepsilon^2 |K(\jmath\alpha \tan\theta))|^2}. \tag{2.143}$$

Thus the passband edge at $\omega = 1$ is mapped to $\omega = \theta_1$ as

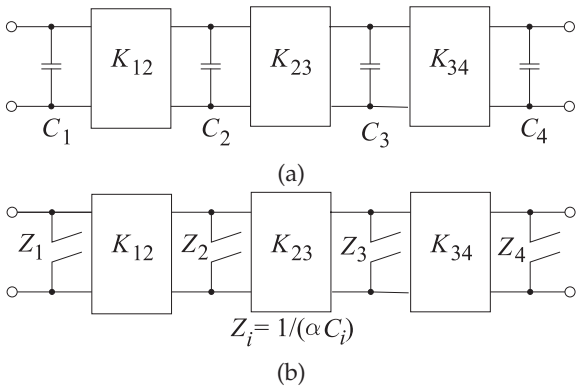
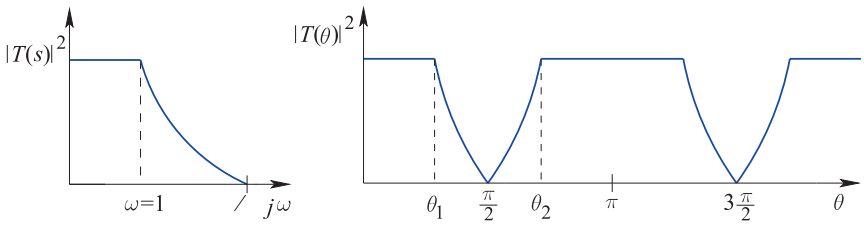
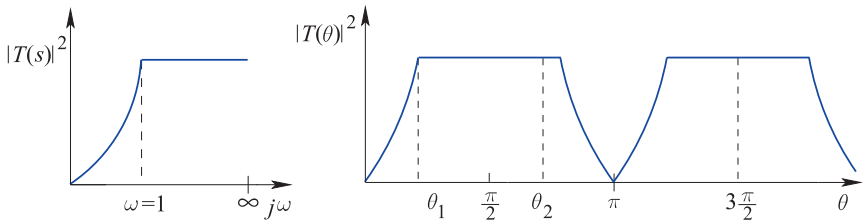


Figure 2-42: Lowpass prototypes: (a) lowpass prototype as a ladder filter with inverters; and (b) lowpass distributed prototype with open-circuited stubs (the impedance looking into the stub is indicated).



(a) Lowpass filter response (b) Distributed lowpass filter response

Figure 2-43: Lowpass to distributed lowpass transformation. $s = j\omega$.



(a) Highpass filter response (b) Distributed highpass filter response

Figure 2-44: Highpass to distributed highpass transformation. $s = j\omega$.

$$\omega = 1 \rightarrow \alpha \tan(\theta_1) \qquad (2.144) \qquad \text{so that} \qquad \alpha = \frac{1}{\tan(\theta_1)}. \qquad (2.145)$$

Recalling that a capacitor is transformed into an open-circuited stub (see Equation (2.139)), Richards's transformation applied to the lowpass filter prototype results in a filter with transmission line elements only, as shown Figure 2-42, provided that the inverters are realized using transmission lines.

The implementation of a lowpass filter in distributed form results in passbands and stopbands that repeat in frequency, as shown in Figure 2-43. This occurs because transmission lines are used to realize lumped elements and the two-port parameters of a transmission line repeat every wavelength (or one-half wavelength in some cases). For example, the input impedance of a stub is the same whether it is one-half wavelength long or one wavelength long.

2.12.4 Richards's Transformation Applied to a Highpass Filter

With reference to Figure 2-44(a), the passband edge at $\omega = 1$ is mapped to θ_1, as shown in Figure 2-44(b). This means that the passband at $\omega = 1$ is mapped

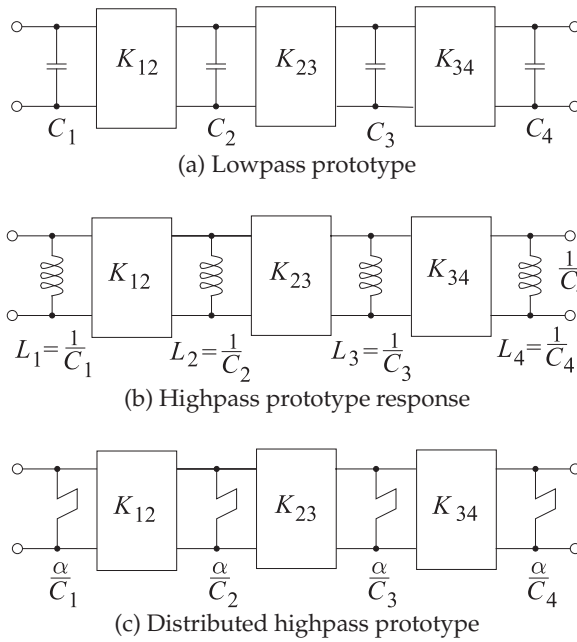


Figure 2-45: Ladder prototype transformation. (The input impedances of the stubs are indicated in (c).)

to θ_1 as (note that θ_1 is the electrical length at the band edge)

$$\omega = 1 \rightarrow \alpha \tan(\theta_1) \qquad (2.146) \qquad \text{so that} \qquad \alpha = \frac{1}{\tan(\theta_1)}. \qquad (2.147)$$

The sequence of steps transforming a lumped lowpass prototype to its distributed highpass prototype are shown in Figure 2-45. As previously discussed, all the inverters can be approximated by transmission lines of length $\pi/2$ at the corner frequency of the filter.

2.13 Kuroda's and Norton's Network Identities

Kuroda's and Norton's network identities are a number of pairs of equivalent networks that facilitate the transformation from one prototype to another, particularly transformations that enable realizable transmission line implementations.

2.13.1 Kuroda's Identities

Kuroda's identities embody a number of specific manipulations using impedance or admittance inverters. They are particularly useful in implementing Richards's transformations as they physically separate transmission line stubs, transform series stubs into shunt stubs, and can change characteristic impedances that are either too small or too high to practically realizable values. Kuroda's identities are a number of equivalent two-port networks, as shown in Figure 2-46. The proof is derived by obtaining the $ABCD$ parameters of the two ports similar to the technique used throughout this chapter. The identities shown in Figure 2-46 are narrowband. With stubs replacing the lumped elements, Kuroda's identities then have broader bandwidths. Kuroda's identities with stubs are shown in Figure 2-47. The major use of these identities is to transform designs with series stubs (in addition to possible shunt stubs) into designs with shunt

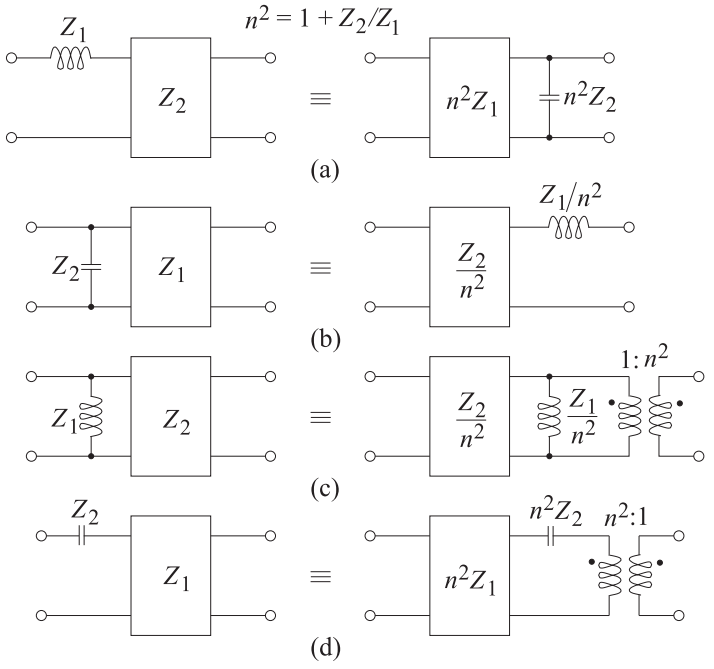


Figure 2-46: Kuroda's identities. Here the inverters are impedance inverters and the designation refers to the impedance of the inverter. Recall that an inverter of impedance Z_1 can be realized by a one-quarter wavelength long transmission line of characteristic impedance Z_1. (As usual, element impedances are indicated.)

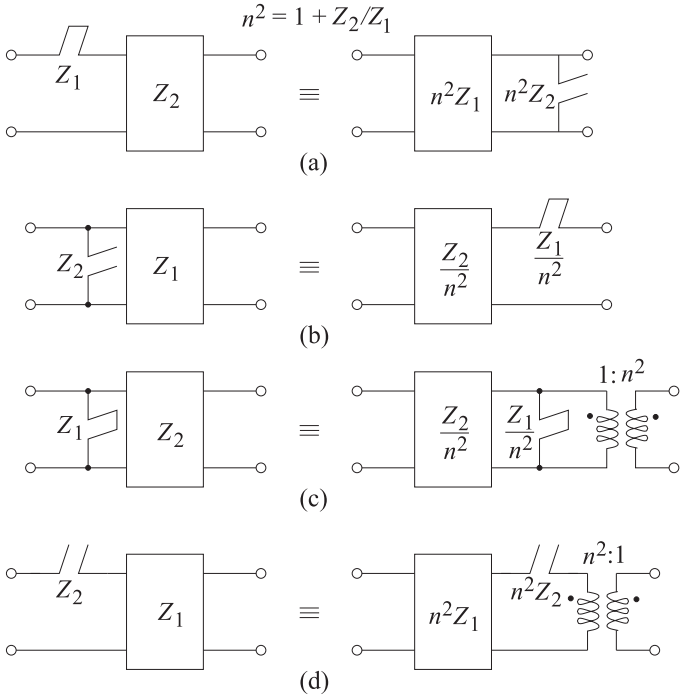


Figure 2-47: The stub form of Kuroda's identities with impedance inverters. The stub impedances shown are the input impedances of the stubs.

stubs only.

To see how these identities are used, consider the identity shown in Figure 2-46(a). The network on the left has a series inductor that, using transmission lines, is realized by a series stub. A series stub cannot be realized in most transmission line technologies, including microstrip. Using the identity shown on the right in Figure 2-46(a), the series stub is replaced by the shunt stub used to realize the shunt capacitor. At the same time, impedance scaling

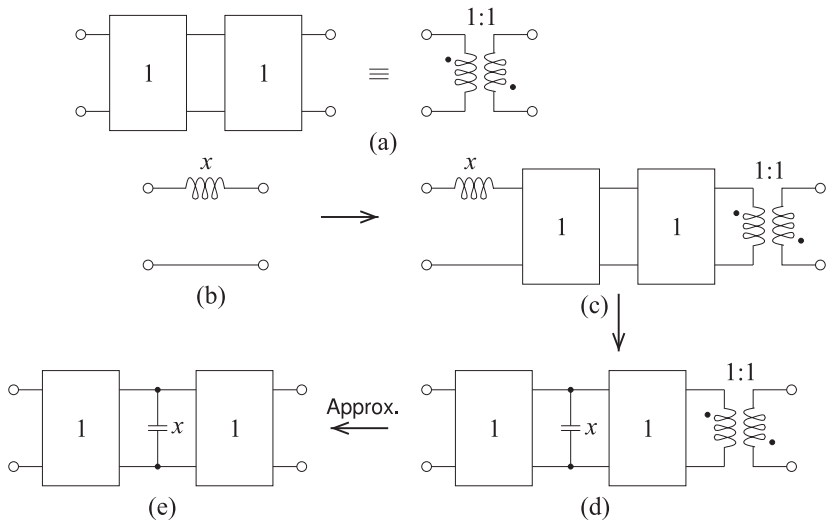


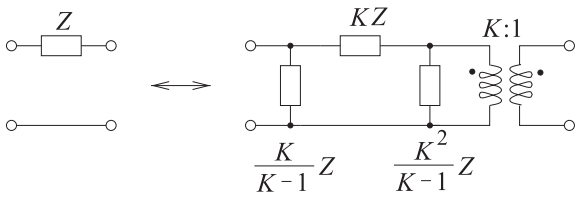
Figure 2-48: Transformation of a series inductor into a shunt capacitor between inverters.

can be used. If the impedance of the inverter in the network on the left is too low, then it can be scaled by a factor n^2, where $n^2 = 1 + Z_2/Z_1$.

One use of Kuroda's transforms is to convert a series inductor into a shunt capacitor. Consider the inductor transformation shown in Figure 2-48. Figure 2-48(a) shows that two inverters in cascade is electrically equivalent to an inverting transformer. Since each inverter is a unity inverter (either a $1\ \Omega$ impedance inverter or a 1 S admittance inverter), the equivalence is a unitary inverting transformer that corresponds to a $180°$ phase change. So the transformation of a series inductor begins with the transformation of the inductor (with value x) in Figure 2-48(b) into the network of Figure 2-48(c) in which the inverter cascade and the inverting transformer cancel each other out. Using the Kuroda identity in Figure 2-47(a), the network in Figure 2-48(c) converts to the electrically identical network in Figure 2-48(d). Since the $180°$ rotation usually does not matter in circuits, the final transformation shown in Figure 2-48(e) is usually acceptable. (The rotation would only matter if there was another path between the input and output as then phasing would affect the way signals on multiple paths combined.) Now the capacitor in Figure 2-48(e) has the numerical value x, the same as the numerical value of the original inductor. This is a result of using unitary inverters. The key result here is that a series inductor is equivalent to a shunt capacitor flanked by two inverters.

2.13.2 Norton's Identities

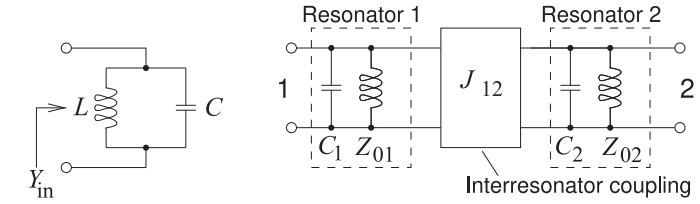
Norton's transformations enable the magnitude of lumped element values to be scaled [16, 17]. Norton's transformations are shown in Figure 2-49 where K is the scaling factor. Additional elements are introduced in these transformations including possibly negatively valued elements. Generally these elements can be combined with other elements so that the elements to be realized are positive (e.g. positive capacitor) and the transformer can also be replaced through subsequent transformations using, for example, Kuroda's identities.



(a) Type 1 Norton's transformation

(b) Type 2 Norton's transformation

Figure 2-49: Norton's transformations.



(a) Bandpass resonator

(b) Coupled bandpass resonators

Figure 2-50: Bandpass resonators.

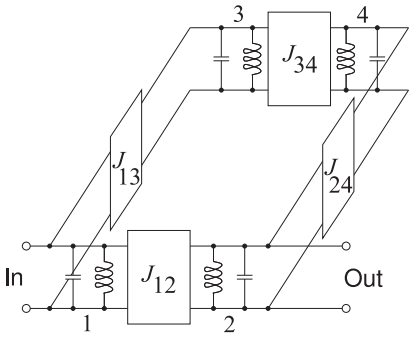


Figure 2-51: The electrical design of a cross-coupled filter. There are four resonators labeled 1, 2, 3, and 4. Thebandpass resonators are coupled by admittance inverters.

2.14 Inter-resonator Coupled Bandpass Filters

The synthesis method that has been described so far leads to bandpass filters that are a cascade of coupled bandpass resonators. A bandpass filter then comprises pairs of resonators of the type shown in Figure 2-50(a) that are coupled using what is called inter-resonator coupling (see Figure 2-50(b)). A generalization of this architecture is that a bandpass filter comprises coupled resonators that do not need to be in a linear cascade. Coupling the output of one resonator latter in a cascade to the input of an earlier resonator is called cross coupling, and a filter that uses this arrangement is said to be a cross-coupled resonator filter or a cross-coupled filter. By reusing resonators in this way, fewer resonators are required to achieve a specified bandpass response and the filter can have higher performance or be smaller. An example of the electrical design of a cross-coupled filter is shown in Figure 2-51.

Design of a cross-coupled bandpass filter is not as systematic as the design of a ladder filter. The design concept is based on using a number of resonators where each resonator on its own has the same resonant

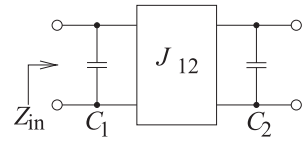


Figure 2-52: Lowpass filter approximation with capacitors separated by an admittance inverter.

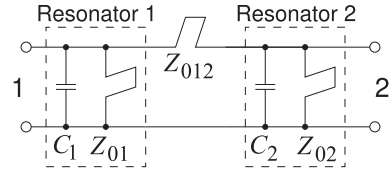


Figure 2-53: Segment of a filter with two resonators. $C_1 = C_2 = 2.7171$ pF, $Z_{01} = Z_{02} = 67.7535\ \Omega$, and $Z_{012} = 442.3836\ \Omega$. The stubs are resonant at $f_r = 2f_0$ where the filter center frequency $f_0 = 1$ GHz. Thus the stubs are $\lambda/8$ long at f_0.

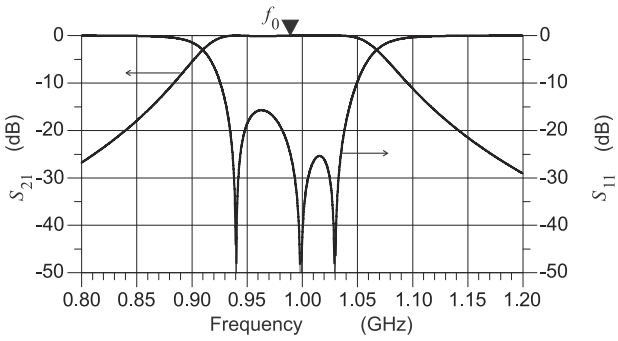
frequency f_0, where f_0 is the center frequency of the filter. The desired filter characteristics are obtained by adjusting the inter-resonator coupling, that is, the coupling between pairs of resonators. Design is not a synthesis process based on a series of mathematical transformations, but instead is based on the understanding of the impact of changing the coupling. The design process will now be described for a pair of resonators.

Coupling of a Pair of Resonators

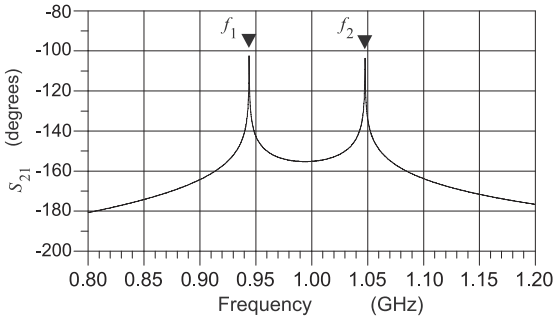
A segment of a bandpass filter fundamentally comprises a pair of coupled bandpass resonators. This is based on the lowpass prototype with a pair of capacitors separated by an admittance inverter (see Figure 2-52). In transforming this to a bandpass filter with a center frequency of f_0, each capacitor is replaced by a resonator such as that shown in Figure 2-50(a). This leads to a pair of resonators, each resonant at f_0, that are coupled by an inverter (see Figure 2-50(b)). The value of the inverter determines the level of coupling of the resonators.

Consider the circuit in Figure 2-53, which consists of a pair of coupled resonators, with each resonator resonant at $f_0 = 1$ GHz, and the resonators are coupled by the series stub with characteristic impedance Z_{012} (i.e., the Z_{012} stub). The circuit of Figure 2-53 is a segment of a Chebyshev 1 GHz bandpass filter with a fractional bandwidth of 10%. This is an interim stage of a bandpass filter design[3] and the system impedance of this circuit segment is $139.4\ \Omega$. The magnitude and phase of the 139.4-Ω S_{21} and S_{11} parameters of this circuit are shown in Figure 2-54(a), and the phase of S_{21} is shown in Figure 2-54(b). Specific features to identify here are the two peaks in the phase response at f_1 and f_2. These are also the frequencies of peak transmission response where $|S_{21}| = 1$. The bandwidth defined by f_1 and f_2 is called the coupling bandwidth ($= f_2 - f_1$) or the bandwidth of the filter (note that this is not the 3-dB bandwidth of the filter). The level of coupling determines the coupling bandwidth. That is, changing the level of coupling provided by the Z_{012} stub changes the position of f_1 and f_2. An interpretation of what is happening is that by coupling the two resonators, the resonate frequency of the resonators on their own, f_0, has been split by the coupling, resulting in two peaks at f_1 and f_2. The higher the coupling, the larger the coupling bandwidth, and the lower the coupling, the smaller the coupling bandwidth. For example, a narrowband bandpass filter comprises

[3] Specifically these resonators are the left-most resonators in a design developed in Chapter 3 (see Figure 3-31).



(a) Magnitude of transmission and reflection responses



(b) Phase of transmission response

Figure 2-54: Response of the coupled resonator network in Figure 2-53.

resonators that have low-level coupling.

So following the choice of an appropriate architecture, that is, the number of resonators and the coupling arrangement, design and manual optimization becomes one of changing the level of coupling between pairs of resonators. Design is not as clean as for ladder filters, but the advantages of a design with fewer resonators is a compelling justification. Invariably, extensive (perhaps a day's worth) manual adjustment of the coupling of the final fabricated filter is required. Overall this is an approach worth adopting for high-value filters such as those used in a basestation, where the high unit costs resulting from manual tuning of individual filters can be justified. This type of filter design is considered further in [3, 18–21].

Relationship of Ladder Synthesis and Inter-resonator Coupling

The relationship of design focused on interresonator coupling and design based on the ladder synthesis approach can be demonstrated for a pair of resonators as follows. Consider the lowpass prototype circuit of Figure 2-52. The input impedance of this circuit is

$$Z_{\text{in}}(s) = \left[sC_1 + J_{12}^2/(sC_2)\right]^{-1}. \tag{2.148}$$

With $s = \jmath\omega$, the poles of Z_{in} occur when

$$J_{12}^2 - \omega^2 C_1 C_2 = 0, \tag{2.149}$$

thus the radian frequencies of the poles are

$$\omega = \pm\frac{J_{12}}{\sqrt{C_1 C_2}}. \tag{2.150}$$

Applying the bandpass transformation (so that the capacitors are replaced by the resonators of Figure 2-50(a)) results in the bandpass circuit of Figure 2-50(b). With this transformation the frequency is transformed as

$$\omega \to \alpha \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right). \tag{2.151}$$

That is, the transmission coefficient of the bandpass filter is

$$T_{\mathrm{BPF}}(\omega) = T_{\mathrm{LPF}} \left(\alpha \left[\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right] \right), \tag{2.152}$$

where $T_{\mathrm{LPF}}(\omega)$ is the transmission coefficient of the lowpass filter at the radian frequency ω. The poles of the bandpass filter are at ω_1, where

$$\alpha \left(\frac{\omega_1}{\omega_0} - \frac{\omega_0}{\omega_1} \right) = + \frac{J_{12}}{\sqrt{C_1 C_2}}, \tag{2.153}$$

and ω_2, where

$$\alpha \left(\frac{\omega_2}{\omega_o} - \frac{\omega_0}{\omega_2} \right) = - \frac{J_{12}}{\sqrt{C_1 C_2}}. \tag{2.154}$$

Then the coupling bandwidth of the filter is

$$f_1 - f_2 = \frac{\omega_1 - \omega_2}{2\pi} = \frac{J_{12}\omega_0}{2\pi\alpha\sqrt{C_1 C_2}}. \tag{2.155}$$

Thus the coupling bandwidth can be directly related to the steps in the ladder-based synthesis of a bandpass ladder filter.

Group Delay

Group delay of a network, and in particular of a filter, is the delay to send information through a network. It is the time required for a modulated carrier signal to appear at the output of a filter after being applied to the input of the filter. Phase delay is a similar measure of delay but does not describe the time it takes to send information. It describes the delay to send a particular phase of a single sinewave. Group delay is a steady-sate concept and so only approximately captures the transient response of a filter.

If a two-port has the transmission coefficient

$$T(s) = S_{21} = a + \jmath b = t \angle \varphi, \tag{2.156}$$

where a and b are its real and imaginary parts, t is its magnitude, and φ is its phase. With $s = \jmath\omega$, the phase of T is

$$\varphi(\omega) = \tan^{-1}\left(\frac{b}{a} \right). \tag{2.157}$$

The group delay, τ_D, is the negative of the derivative of this phase as follows:

$$\text{Group delay} = \tau_g(\omega) = -\frac{d\varphi}{d\omega} = -\frac{d}{d\omega}\left[\tan^{-1}\left(\frac{b}{a} \right) \right]. \tag{2.158}$$

This compares to the phase delay:

$$\text{Phase delay} = \tau_\varphi(\omega) = -\frac{\varphi}{\omega} = -\tan^{-1}\left(\frac{b}{a} \right). \tag{2.159}$$

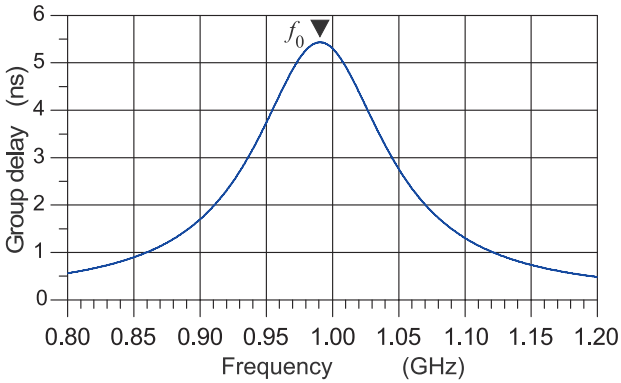
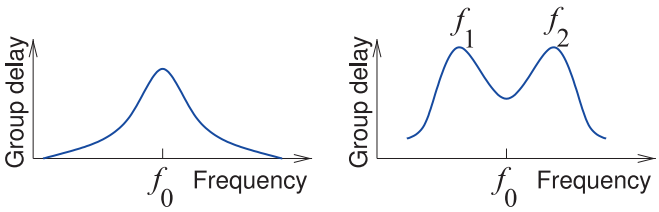


Figure 2-55: Group delay of a resonator.



(a) Single resonator (b) Coupled resonator pair

Figure 2-56: Group delay of resonators in a filter. Additional delay is introduced by the coupling inverter.

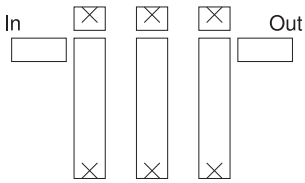


Figure 2-57: Microstrip layout of a third-order combline filter based on coupled lines. Each of the vertical microstrip line forms a resonator with the top gap capacitors. The input and output gap capacitors provide impedance matching.

Figure 2-55(c) shows the group delay response of one of the resonators in Figure 2-53, where the group delay peaks at the resonant frequency of the resonator f_0. The peak of the group delay occurs at the resonant frequency of the resonator without loading. From examination of the phase of S_{21} (see Figure 2-54(b)), it is clear that the group delay peaks very close to f_1 and f_2. So the coupled resonator pair has two peaks in the group delay. This can be seen in Figure 2-56, which plots the group delay of a single resonator and a coupled pair of resonators.

2.15 Bandpass Filter Topologies

The essential structure of a bandpass filter comprises resonators that are coupled to each other. A large number of filter architectures based on this concept have been deployed. A parallel coupled line (PCL) filter called a combline filter is shown in Figure 2-57. Here the microstrip lines form the resonator and the coupling of parallel lines provides the interresonator coupling. A variation on a filter with PCL resonators is shown in Figure 2-58 [22]. This filter uses edge-coupled microstrip resonators that are $\lambda/2$ long (at the center frequency). Figure 2-59 shows another distributed bandpass filter topology utilizing end-coupled microstrip resonators. That is, the gaps provide the interresonator coupling. All transmission line bandpass filters have spurious passbands [23–25]. The root cause of the spurious responses derives from the transformation of the parallel LC resonators into their

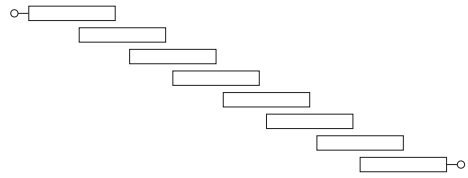


Figure 2-58: Microstrip layout of a parallel coupled bandpass filter.

Figure 2-59: General microstrip layout for an end-coupled bandpass filter (series coupling gaps between cascaded straight resonator elements).

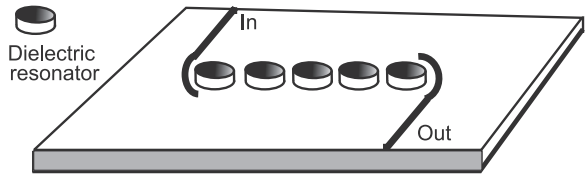
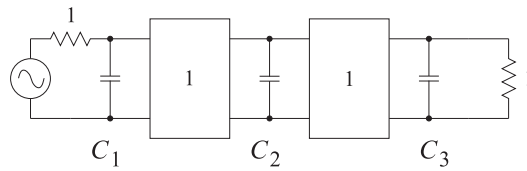


Figure 2-60: A microstrip coupled dielectric resonator bandpass filter configuration.



$C_1 = 0.85239\ \text{F} = C_3 \qquad C_2 = 1.10422\ \text{F}$

Figure 2-61: Bandpass filter approximation.

transmission line form.

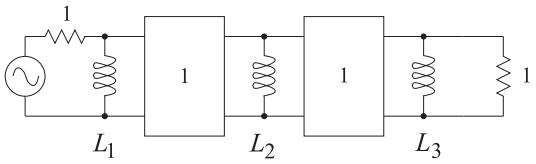
A **dielectric-resonator**-based bandpass filter is shown in Figure 2-60 [26]. Here the pucks are the bandpass resonators and they are coupled and the evanescent fields outside the pucks provide the interresonator coupling. The pucks shown are cylinders of high-permittivity material (typically having a relative permittivity of 500–85 000) with an approximate magnetic wall at the cylindrical surface of the puck. Thus the puck resonates when its diameter is approximately $\lambda/2$.[4]

With all these filters the interresonator coupling functions as an inverter. Thus the basic functional unit of the bandpass filters is the coupled resonator structure shown in Figure 2-50(b).

2.16 Case Study: Design of a Bandstop Filter

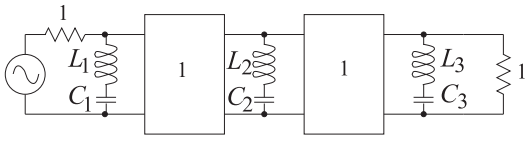
The design of a bandstop filter begins with the lowpass filter prototype shown in Figure 2-61. To the lowpass prototype, the highpass transformation is applied to obtain the highpass prototype of Figure 2-62. Picking a center frequency of approximately 1 GHz and corner frequencies of $f_1 = 950$ MHz and $f_2 = 1050$ MHz, corresponding to a bandwidth of approximately 10%, the bandstop transformation is now applied. This results in the prototype of Figure 2-63. Finally, scaling the system impedance to $50\ \Omega$ leads to the prototype of Figure 2-64. The impedance inverters must remain set at $50\ \Omega$ in order to obtain a broad match. In the passband, energy will pass at all

[4] A better estimate is developed from the zeros of Bessel functions, as the fields inside the pucks have a Bessel function dependence (this is the form of the solution of the wave equation in cylindrical coordinates).



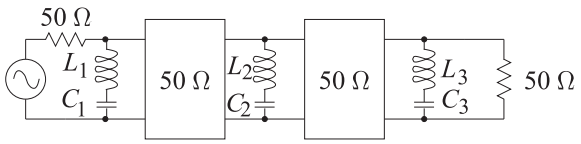
$L_1 = L_3 = \frac{1}{C_1} = \frac{1}{C_3} = 1.1743\ \text{H}$ $\quad L_2 = \frac{1}{C_2} = 0.9064\ \text{H}$

Figure 2-62: Bandstop filter prototype.



$L_1 = L_3 = 1.868\ \text{nH}$ $\quad C_1 = C_3 = 13.587\ \text{pF}$
$L_2 = 1.442\ \text{nH}$ $\quad C_2 = 17.601\ \text{pF}$

Figure 2-63: Bandstop filter following transformation from the lowpass prototype.



$L_1' = L_3' = 1.868 \times 50 = 93.448\ \text{nH}$
$C_1' = C_3' = 13.587/50 = 0.2717\ \text{pF}$
$L_2' = 1.442 \times 50 = 72.1362\ \text{nH}$
$C_2' = 17.601/50 = 0.352\ \text{pF}$

Figure 2-64: Bandstop filter after impedance transformation.

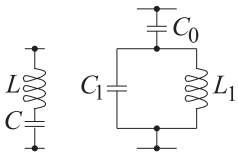


Figure 2-65: Transformations of the resonators in the bandstop filter to obtain realizable values. The series LC resonator in (a) is transformed to the form in (b).

frequencies.

In Figure 2-64, the inductor values are relatively large and the capacitor values are relatively small so that it will be difficult to realize the filter in either lumped or distributed forms. These values must be scaled to obtain realizable values. One possible transformation is shown in Figure 2-65. To establish that the left-hand and right-hand networks are equivalent, at least near one frequency, the impedances and derivatives must be matched. For the circuit in Figure 2-65(a),

$$Z_1 = \jmath\frac{\omega^2 LC - 1}{\omega C} \qquad (2.160) \quad \text{and} \quad \frac{dZ_1}{d\omega} = \jmath\frac{\omega^2 LC + 1}{\omega^2 C}, \qquad (2.161)$$

and for the circuit in Figure 2-65(b),

$$Z_2 = \jmath\frac{\omega^2 L_1 C_1 - 1 + \omega^2 L_1 C_0}{\omega C_0\left(1 - \omega^2 L_1 C_1\right)} \qquad (2.162)$$

and

$$\frac{dZ_2}{d\omega} = \jmath\frac{\omega^4 L_1^2 C_1^2 - 2\omega^2 L_1 C_1 + \omega^4 L_1^2 C_0 C_1 + \omega^2 L_1 C_0 + 1}{\omega^2 C_0\left(\omega^2 L_1 C_1 - 1\right)^2}. \qquad (2.163)$$

Equating the above enables C_0 and C_1 to be found for a chosen value of L_1. Thus the series LC resonators in Figure 2-66(a and c) are replaced by the

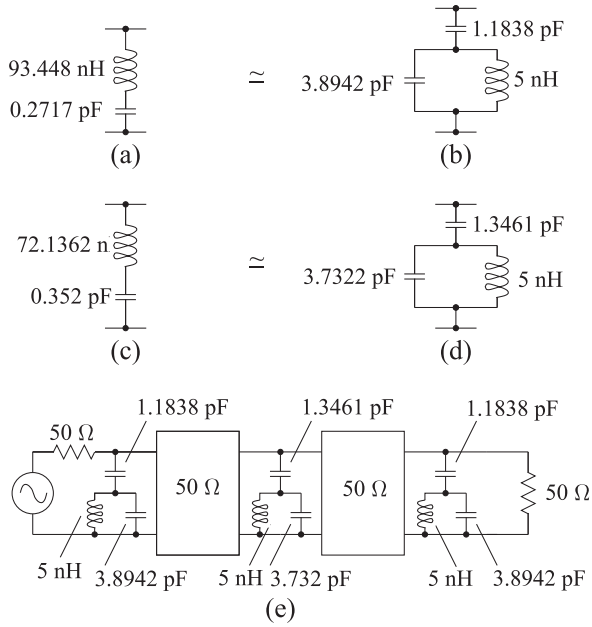


Figure 2-66: Intermediate bandstop filter prototype.

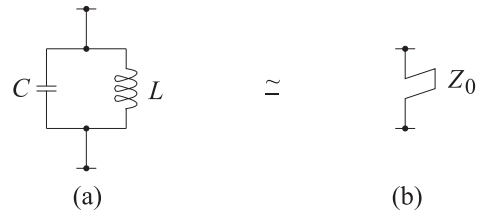


Figure 2-67: Equivalence of shunt bandpass resonator to shunt short-circuited stub.

networks in Figure 2-66(b and d). This results in the filter of Figure 2-66(e).

At this stage the bandpass resonators are then equated to short-circuited stubs by equating the admittance, Y_1, of the lumped circuit in Figure 2-67(a) with the admittance, Y_2, of the stub in Figure 2-67(b). That is, by equating

$$Y_1 = \jmath\frac{\left(\omega^2 CL - 1\right)}{\omega L} \quad \text{and} \quad Y_2 = \frac{1}{\jmath Z_0 \tan\left(\frac{\pi}{2}\frac{\omega}{\omega_r}\right)}. \tag{2.164}$$

The characteristic impedance of the stub, Z_0, is selected so that the frequency, ω_r, is not too far above the upper band-edge frequency of the filter, in this case 1.05 GHz. Choosing $Z_0 = 20\ \Omega$ results in the stub transformations shown in Figure 2-68(a–d). The bandstop filter prototype with stubs is shown in Figure 2-68(e). The final physical layout of the bandstop filter is shown in Figure 2-69. The response of the final bandstop filter design is shown in Figure 2-70.

2.17 Active Filters

With active circuits, losses in lumped-element components can be compensated for by the gain of active devices resulting in filters with high Q. Inverters can also be realized efficiently. The net result is that active filters provide a performance per unit area advantage over passive element realizations. Of course, active filters are limited to low-level signals (such as in the receive circuitry, and prior to the power amplifier stages) as the full power of the

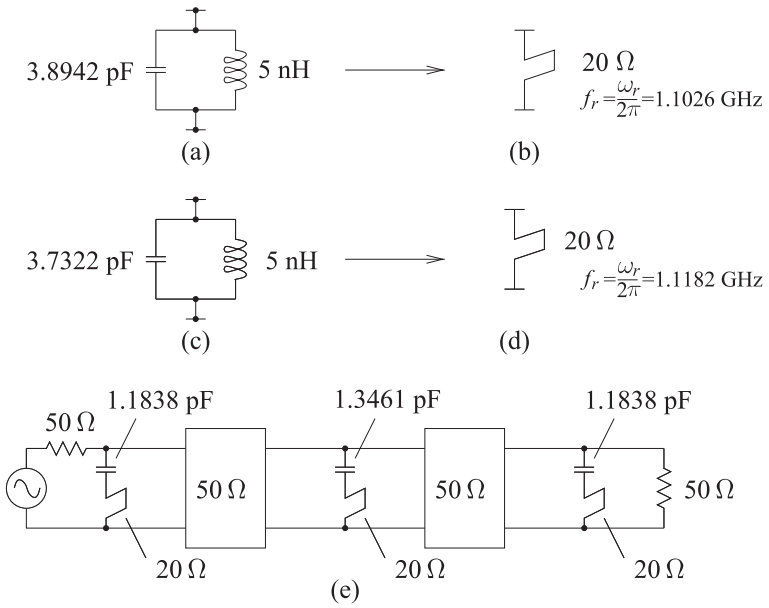


Figure 2-68: Bandstop filter prototype using stub approximations. The stub in (b) is the transmission line approximation of the parallel resonant circuit in (a). The stub in (d) approximates the circuit in (c).

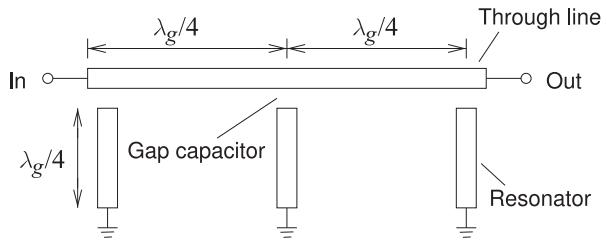


Figure 2-69: Physical layout of a bandstop filter in microstrip.

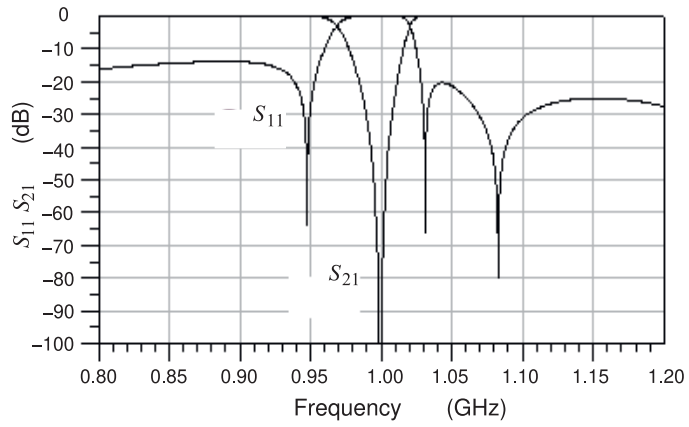


Figure 2-70: Response of the bandstop filter shown in Figure 2-69.

signal must be handled by the active devices in the active filter. With active devices, additional noise is introduced and so the noise figure of an active filter must be considered at RF where bandwidths are significant.

At microwave frequencies active filters based on traditional low-frequency concepts have relatively low Q. Higher Q and narrower-band applications require active inductors or distributed techniques. The smallest active filters are the ones that use active inductors and the largest use distributed transmission line elements.

Sometimes the main role of a filter is to limit the dynamic range of signals presented to active circuits, thereby limiting distortion. Clearly active filters are not suitable when large out-of-band signals are a concern.

2.17.1 Radio Frequency Active Filters

There are several problems with lumped-element filters. High-order filters require many poles and zeros that must be precisely placed, requiring small tolerances. With active filters it is possible, to some degree, to achieve higher-order filters with fewer reactive elements and so the tolerancing problem is considerably reduced. Also, there are a few designs that can be tuned electronically. Lumped-element bandpass filters and highpass filters require inductors. Inductors are a major problem, as RF inductors are lossy and also only relatively small values can be obtained. On-chip inductors take considerable area and can thereby dominate the cost of RFICs. With active filters, feedback and capacitors can be used to realize inductor-like characteristics, defined as a positive imaginary impedance increasing with frequency at least over a small frequency range. Many times, and especially at lower radio frequencies, lumped inductors are no longer required. The size of capacitors in filters is another issue, and using feedback and gain, the effective value of capacitors can be magnified. All this comes with limits. The principle limit is that the cutoff frequency of the transistors must be much higher than the operating frequency of the filters. This is a cost incurred whenever feedback is used. A rough rule of thumb is that the transistor cutoff frequency should be about 10 times the operating frequency. So for a 5 GHz bandpass filter, a 50 GHz transistor process is required. The other significant cost is that active filters are only suitable for small signals. The efficiency, a measure of the amount of DC power used in producing the RF output signal, is then not a consideration. Active filters cannot be used where signal levels are likely to be large (such as at the output of a power amplifier or at the receiver input, due to likely out-of-band signals).

The essential aspect of low-RF active filters is the use of feedback, with reactive feedback elements, to realize a frequency-domain transfer function with poles and zeros. The aim is not to synthesize effective capacitors and inductors, but to focus on the overall response. Active filters are generally developed as stages, with each stage realizing a second- or higher-order transfer function. Design is simulation driven, as this is the only way to account for the active device parasitics.

In operational amplifier design, feedback is used to achieve gain stability, but at the cost of reduced bandwidth. If reactive feedback elements are used, high-order transfer functions can be obtained with just a few reactive elements. In Figure 2-71, a twin-T notch network is connected in the operational amplifier feedback path to obtain a bandpass filter [27]. Away from the notch frequency the feedback path has a high impedance and the overall amplifier gain falls off. As the signal frequency approaches the notch frequency, the feedback path becomes effective and the amplifier gain increases. The notch frequency, f_n, is proportional to the RC product of the feedback components and the Q is proportional to the amplifier gain. The governing equations for this filter are

$$f_n = 1/(2\pi RC) \qquad \text{and} \qquad Q = (1+G)/4. \tag{2.165}$$

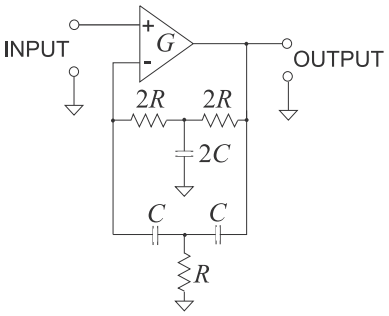
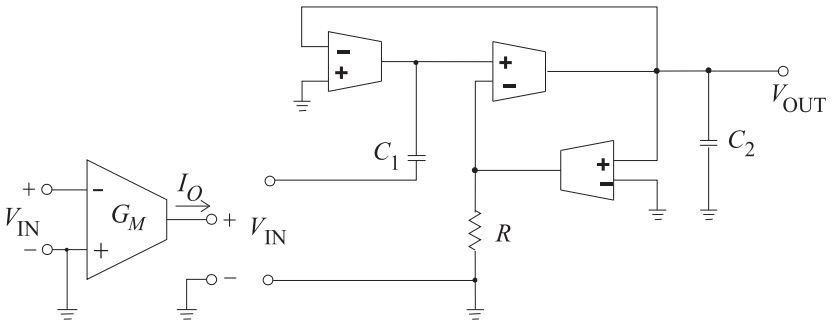


Figure 2-71: Active twin-T bandpass filter: center frequency $f_0 = \pi RC/4$ and $Q = (G+1)/4$.



(a) Schematic (b) RF biquad bandpass filter

Figure 2-72: Operational transconductance amplifier with transconductance G_M. In (a) $I_O = G_M V_{IN}$.

2.17.2 Biquadratic Filters

Biquadratic filters are an important class of filters and are the basic building block used in analog and RF active filters. Biquad filters commonly use operational **transconductance amplifiers** (**OTAs**), the schematic of which is shown in Figure 2-72(a). Here $I_O = G_M V_{IN}$, which is also the basic characteristic of an individual transistor. These circuits have characteristics typical of operational amplifiers, including gain stability and resistance to component tolerance variations. OTA-based filters have the advantage of easy tunability, with pole and zero frequencies electronically adjustable. An RF bandpass **biquad** filter is shown in Figure 2-72(b). Stages such as this are cascaded to realize high-order filters.

Biquadratic filters implement the transfer function

$$H(s) = \frac{V_{OUT}}{V_{IN}} = \frac{N(s)}{D(s)} = \frac{a_2 s^2 + a_1 s + a_0}{s^2 + b_1 s + b_0} = \frac{a_2(s+z_1)(s+z_2)}{(s+p_1)(s+p_2)}. \quad (2.166)$$

This is referred to as a biquadratic function and filters that implement this function are called biquadratic filters or simply biquad filters. There are several special types of biquad filters.

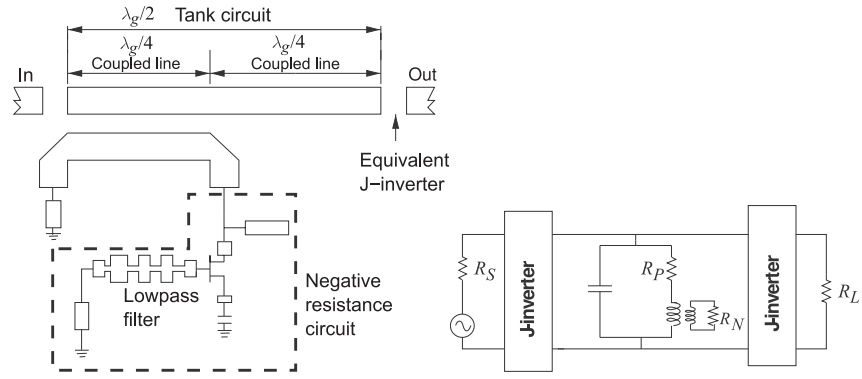
The form of the biquadratic function for a second-order lowpass filter is

$$H_{LP}(s) = \frac{a_0}{s^2 + b_1 s + b_0} = \frac{K\omega_p^2}{s^2 + (\omega_p/Q_p)s + \omega_p^2}. \quad (2.167)$$

$H_{LP}(s)$ has a double zero at $s = \infty$ and thus negligible response at very high frequencies.

The highpass form of the biquadratic filter is described by

$$H_{HP}(s) = \frac{a_2 s^2}{s^2 + b_1 s + b_0} = \frac{K s^2}{s^2 + (\omega_p/Q_p)s + \omega_p^2}. \quad (2.168)$$



(a) Layout (b) Equivalent circuit model

Figure 2-73: Active coupled distributed filter. After [30].

The response of $H_{\mathrm{HP}}(s)$ at high frequencies is K and the response at very low frequencies goes to zero.

The bandpass form of the biquadratic filter is described by

$$H_{\mathrm{BP}}(s) = \frac{a_1 s}{s^2 + b_1 s + b_0} = \frac{K(\omega_p/Q_p)s}{s^2 + (\omega_p/Q_p)s + \omega_p^2}. \tag{2.169}$$

The response of $H_{\mathrm{BP}}(s)$ at high and low frequencies goes to zero, and only at and near the center frequency, $\omega = \omega_p$, is there a reasonable response.

The band-reject or notch form of the biquadratic filter is described by

$$H_{\mathrm{BR}}(s) = \frac{a_2 s^2 + a_0}{s^2 + b_1 s + b_0} = \frac{K(s^2 + \omega_z^2)}{s^2 + (\omega_p/Q_p)s + \omega_p^2}. \tag{2.170}$$

The response of $H_{\mathrm{BR}}(s)$ at high and low frequencies is high, but there is a double zero at the notch frequency, $\omega = \omega_z$, where the response is very low.

With all of the biquadratic filters, the sharpness of the response is determined by Q_p. The edge frequency or center frequency is ω_p for the bandpass, lowpass, and highpass filters, and the notch frequency is ω_z for the bandstop filter.

2.17.3 Distributed Active Filters

At high microwave frequencies, good results can be obtained by combining distributed elements with a gain stage. Commonly the active devices are used as coupling devices. A negative resistance can be produced using the active devices that can compensate for the losses associated with the transmission line elements or lumped passive elements in the filter. The concept uses a resonant tank circuit formed by a transmission line or lumped-element resonator with a negative resistance coupled into the tank circuit by the active devices [28–30]. This concept is illustrated in Figure 2-73. Here a one-half wavelength long resonator forms a resonant circuit (commonly called a tank circuit in this context). A negative resistance is coupled into the tank circuit to compensate for losses in the resonator and the result is effectively a lossless tank circuit.

An active resonator circuit is shown in Figure 2-74. The design of the lossless resonator is based on the negative resistance obtained when the capacitor is connected at the source of the MESFET. This element in the

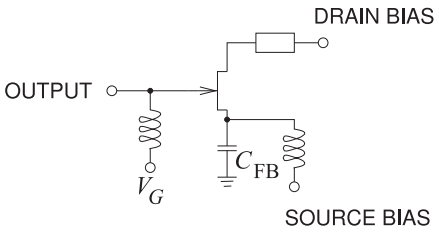


Figure 2-74: An active resonator circuit. After [31].

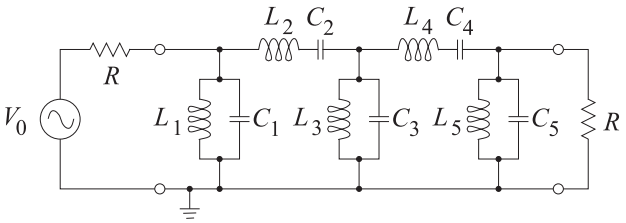


Figure 2-75: Lumped-element 5th-order Chebyshev filter.

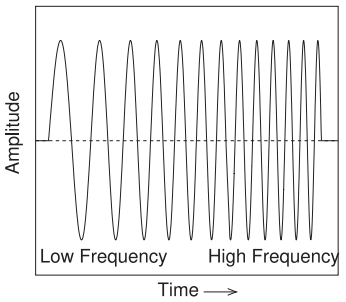


Figure 2-76: A linear chirp.

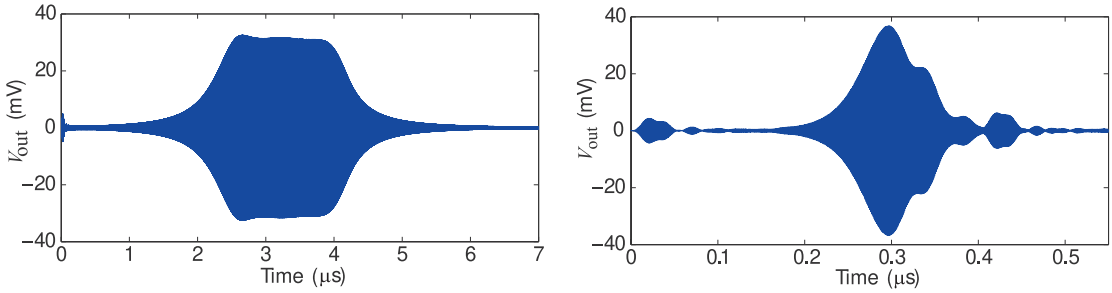
source provides a feedback path between the output of the circuit, the drain-to-ground voltage, and the input gate-to-source voltage. An increase in the drain-source current leads to a voltage at the source that changes the gate-source voltage. This induces a negative resistance that is adjusted through the feedback capacitance, C_{FB}, to compensate for inductor losses. An additional inductor, L_P, is added at the gate. The inductor resonates with the series combination of C_{FB} and C_{GS}. C_{FB} can be implemented using a varactor diode to enable electronic tuning.

2.18 Transient Response of a Bandpass Filter

A bandpass filter is normally characterized by its center frequency and bandwidth, and sinusoidal signals that are within the bandwidth are transmitted by the filter and those outside the passband are reflected by the filter. However, if the signal changes frequency quickly, the transient response of the filter can be unexpected [32, 33].

Figure 2-75 is a fifth-order filter with each pair of shunt resonators coupled by a series resonator. In response to an RF pulse (a pulse of sinusoidally varying RF), the filter passes energy in-band when all of the resonators have reached steady state. When the RF pulse is removed, the resonators will loose energy and eventually the RF signals on the resonators disappear. There will be a finite time to "charge" and "discharge" the filter's resonators.

A classic test of the RF response over frequency is to use a linear RF chirp, as shown in Figure 2-76. With a linear **chirp**, the frequency of the signal in the pulse changes linearly and smoothly over time from one frequency to another. Figure 2-77(a) presents the transient response at the output of



(a) Slow chirp response (b) Fast chirp response

Figure 2-77: Output transient response of a 3rd-order Chebyshev filter with a center frequency of 1 GHz and a 30 MHz bandwidth excited by a -20 dBm linear chirp from 950 MHz to 1050 MHz: (a) chirp rate = 20 MHz/μs, (b) chirp rate = 400 MHz/μs. At the slower chirp rate, the filter response is approximately a superposition of steady-state responses as the frequency changes (i.e., the response is quasi-stationary). At the higher chirp rate, the filter response can no longer be approximated as a sum of steady-state responses (i.e., it is no longer quasi-stationary). After [34].

a 30 MHz-bandwidth 1 GHz filter excited by a 1 μs-long linear chirp from 950 MHz to 1050 MHz with a chirp rate of 20 MHz/μs. This is a relatively slow chirp. The chirp starts at 0 μs and stops at 5 μs. Thus the frequency of the chirp is in-band at 1.75 μs and is out-of-band at 3.25 μs. The chirp is sufficiently slow that the transient transmission response corresponds to the frequency response of the filter. However, if the chirp is fast, the resonators may not reach steady state, and then the transient response of the filter cannot be simply extrapolated from its frequency-domain response. This is seen in Figure 2-77(b), which is the response of the filter to a fast chirp with a chirp rate of 400 MHz/μs and a chirp duration of 250 ns, and beginning at 950 MHz and ending at 1050 MHz. The RF chirp begins out of band at 0 ns, becomes in-band at 0.0875 μs, and goes out-of-band again at 0.1625 μs before turning off at 0.25 μs. Interestingly, there is an almost immediate output response, around 0.03 μs when the RF chirp is out of band. Overall the chirped-filter response does not follow the frequency response of the filter.

Another view of the same phenomenon is examined for a 7th-order Chebyshev filter with a center frequency of 900 GHz and 34 MHz bandwidth from 883 MHz to 917 MHz. The frequency-domain transmission response of the filter is shown in Figure 2-78. The transmission responses to 0.25 μs-long RF pulses of different frequencies are shown in Figure 2-79. The frequency of the RF pulse is shown in each subplot. The top-left plot is the response when the RF pulse is in-band. Here the RF pulse is applied at 0 μs and removed at 0.25 μs. An initial delay in the transmission response and what is referred to as a **long-tail response** is seen when the RF pulse is removed. The other plots in Figure 2-79 show the filter response as the RF is backed off from the passband. So, even when the RF is out of band, there can be an appreciable output from the filter at the beginning and end of the RF pulse.

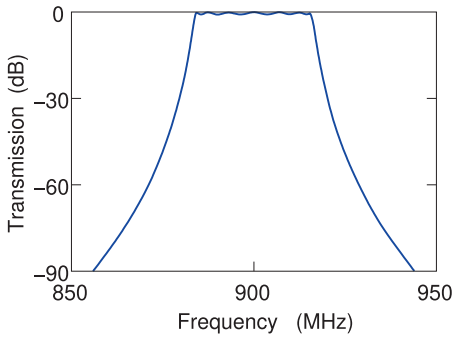


Figure 2-78: Frequency-domain transmission response of a 7th-order Chebyshev filter with a center frequency of 900 GHz and 34 MHz bandwidth from 883 MHz to 917 MHz.

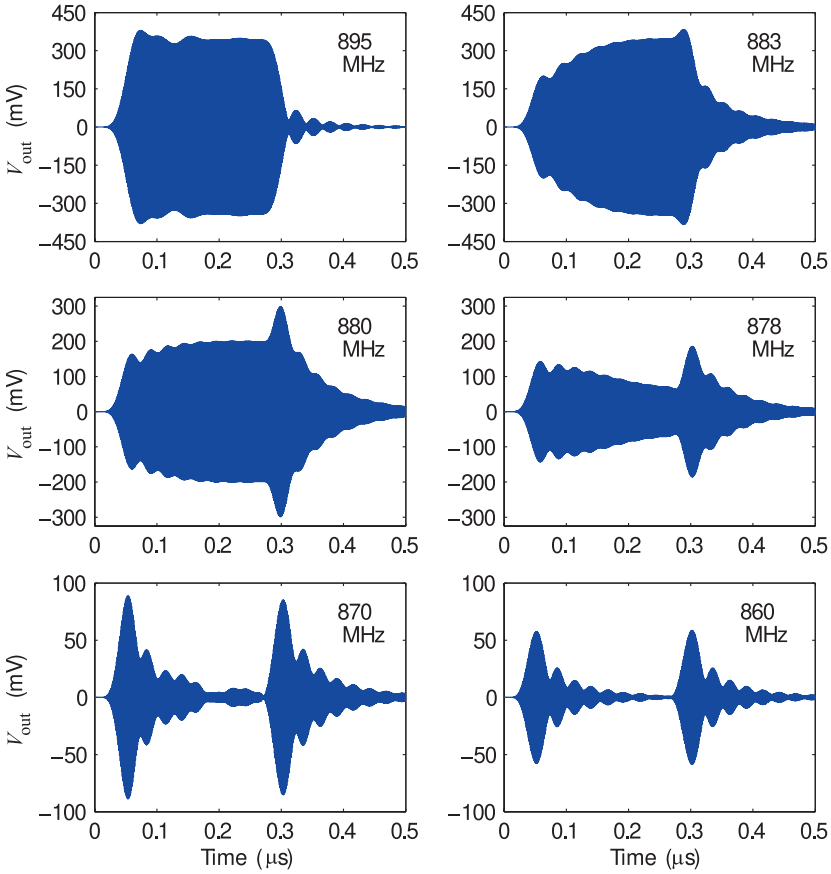


Figure 2-79: Transient response at the output of the 900 MHz 7th-order Chebyshev bandpass filter with a passband from 883 MHz to 917 MHz. The frequency response is shown in Figure 2-78. After [34].

2.19 Summary

Radio frequency and microwave filter design combines the mathematical synthesis of a circuit with the required performance and the intuitive realization that particular structures inherently have a desired frequency selectivity. The art of filter design is to direct the mathematical synthesis and companion circuit so that the circuit structures in the synthesized filter prototype match the functionality of physical microwave structures. A very important step in this process is the development of circuit equivalent models of physical structures. Once a microwave filter with appropriate topology has been designed and the close-to-final physical filter is designed, the physical layout is optimized in a circuit simulator to account for

parasitics and higher-order EM effects.

A similar design procedure applies to active filter designs where particular circuit topologies, for example, the biquad network blocks, are identified and exploited. The synthesis approach provides design insight and exploitation of all parameters of a transfer function, leading to optimum network topologies. Synthesis can be time consuming and specialized, but it is the only way to develop filters with optimum performance. Insight gained during design identifies yield issues and provides the basis for inventing new filter topologies.

Further Reading

The design of microwave filters has a rich tradition. Many excellent books and articles have been written about microwave filter design techniques. Books and articles with extensive treatments are references [1, 3, 4, 14, 16–19, 35–46]. However the most important systematic approaches to microwave filter design were presented in this chapter. Many journal and conference papers present topologies in various technologies that can be used to implement filters. The best way to locate these papers is to search using the specifics of the technology of interest. For example "wideband microstrip bandstop filters on glass substrates" will yield a list of papers on the topic. There are many small companies that specialize in different types of filter design. For very high performance filters, for example in basestation applications, the number of designers and vendors is quite small and filter designers follow the literature and patents closely. For designers building microwave systems with small to medium volumes the best approach is to use microstrip design. For bandpass microstrip design the most common choice is to use filter design based on coupled microstrip design which is considered in the next chapter.

2.20 References

[1] G. Matthaei, L. Young, and E. Jones, *Microwave Filters, Impedance-Matching Networks and Coupling Structures*. McGraw-Hill, 1965, reprinted in 1980, Artech House.

[2] D. M. Pozar, *Microwave engineering*, 4th ed. John Wiley & Sons, 2012.

[3] I. Hunter, *Theory and Design of Microwave Filters*. IEE Press, 2001.

[4] J. Scanlan and R. Levy, *Circuit Theory*. Oliver & Boyd, 1973, vol. 2.

[5] ——, *Circuit Theory*. Oliver & Boyd, 1973, vol. 1.

[6] V. Belevitch, "Techbyshev filters and amplifier networks," *Wireless Engineer*, pp. 106–110, Apr. 1952.

[7] H. Orchard, "Formulae for ladder filters," *Wireless Engineer*, pp. 3–5, Jan. 1953.

[8] R. Daniels, *Approximation Methods for Electronic Filter Design*. McGraw-Hill, 1974.

[9] M. Lutovac, D. Miroslav, D. Tosic, V. Dejan, and B. Evans, *Filter Design for Signal Processing using MATLAB and Mathematica*. Prentice Hall, 2001.

[10] M. Steer, *Microwave and RF Design, Transmission Lines*, 3rd ed. North Carolina State University, 2019.

[11] ——, *Microwave and RF Design, Networks*, 3rd ed. North Carolina State University, 2019.

[12] P. Richards, "Resistor-transmission-line circuits," *Proc. of the IRE*, vol. 36, no. 2, pp. 217–220, Feb. 1948.

[13] ——, "General impedance-function theory," *Quarterly Applied Mathematics*, vol. 6, no. 6, pp. 21–29, 1948.

[14] J. Rhodes, *Theory of Electrical Filters*. John Wiley & Sons, 1976.

[15] H. Baher, *Synthesis of Electrical Networks*. John Wiley & Sons, 1984.

[16] M. E. Van Valkenburg, *Introduction to modern network synthesis*. John Wiley & Sons, 1960.

[17] A. I. Zverev *et al.*, *Handbook of filter synthesis*. John Wiley & Sons, 1967.

[18] R. Cameron, R. Mansour, and C. Kudsia,

Microwave Filters for Communication Systems: Fundamentals, Design and Applications. John Wiley and Sons, 2007.

[19] R. Cameron, "Advanced coupling matrix synthesis techniques for microwave filters," *IEEE Trans. on Microwave Theory and Techniques*, vol. 51, no. 1, pp. 1–10, Jan. 2003.

[20] I. Hunter, J. Rhodes, and V. Dassonville, "Dual-mode filters with conductor-loaded dielectric resonators," *IEEE Trans. on Microwave Theory and Techniques*, vol. 47, no. 12, pp. 2304–2311, Dec. 1999.

[21] W. Fathelbab and M. Steer, "A reconfigurable bandpass filter for RF/microwave multifunctional systems," *IEEE Trans. on Microwave Theory and Techniques*, vol. 53, no. 3, pp. 1111–1116, Mar. 2005.

[22] A. Nakamura, K. Hibino, F. Yamamoto, and S. Kamihashi, "Expert system for microwave filter design," in *1990 IEEE MTT-S Int. Microwave Symp. Dig.*, May 1990, pp. 1183–1186.

[23] W. Fathelbab and M. Steer, "Broadband network design," in *Multifunctional Adaptive Microwave Circuits and Systems*, M. Steer and W. Palmer, Eds., 2008, ch. 8.

[24] ——, "Parallel-coupled line filters with enhanced stopband performance," *IEEE Trans. on Microwave Theory and Techniques*, vol. 53, no. 12, pp. 3774–3781, Dec. 2005.

[25] P. Cheong, S.-W. Fok, and K.-W. Tam, "Miniaturized parallel coupled-line bandpass filter with spurious-response suppression," *IEEE Trans. on Microwave Theory and Techniques*, vol. 53, no. 5, pp. 1810–1816, May 2005.

[26] I. Bahl and P. Bhartia, *Microwave Solid State Circuit Design*. John Wiley & Sons, 1988.

[27] F. Rosenbaum, R. Gregory, W. Richard, W. Ou, F. Kuhns, and T. Trimble, "An MMIC twin-tee active bandpass filter," in *1993 IEEE MTT-S Int. Microwave Symp. Dig.*, 1993, pp. 361–364.

[28] B. Kapilevich, "Active microwave filters," *Telecommunication Radio Engineering*, vol. 4, no. 2, pp. 51–58, Feb. 1985.

[29] C. Chang and T. Itoh, "Narrowband planar microwave active filter," *Electronics Letters*, vol. 25, no. 18, pp. 1228–1229, Aug. 1989.

[30] C.-Y. Chang and T. Itoh, "A varactor-tuned, active microwave band-pass filter," in *1990 IEEE MTT-S Int. Microwave Symp. Dig.*, May 1990, pp. 499–502.

[31] U. Karacaoglu and I. Robertson, "High selectivity varactor-tuned MMIC bandpass filter using lossless active resonators," in *IEEE 1994 Microwave and Millimeter-Wave Monolithic Circuits Symp. Dig. of Papers*, May 1994, pp. 237–240.

[32] G. Mazzaro, M. Steer, and K. Gard, "Intermodulation distortion in narrowband amplifier circuits," *IET Microwaves, Antennas Propagation*, vol. 4, no. 9, pp. 1149–1156, Sep. 2010.

[33] G. Mazzaro, M. Steer, K. Gard, and A. Walker, "Response of RF networks to transient waveforms: Interference in frequency-hopped communications," *IEEE Trans. on Microwave Theory and Techniques*, vol. 56, no. 12, pp. 2808–2814, Dec. 2008.

[34] G. Mazzaro, "Time-frequency effects in wireless communication systems," Ph.D. dissertation, North Carolina State University, 2009.

[35] J. Malherbe, *Microwave Transmission Line Filters*. Artech House, 1979.

[36] J.-S. Hong and M. Lancaster, *Microstrip Filters for RF/Microwave Applications*. John Wiley & Sons, 2001.

[37] H. Howe, *Stripline Circuit Design*. Artech House, 1974.

[38] R. E. Collin, *Foundations for Microwave Engineering*. John Wiley & Sons, 2007.

[39] D. G. Swanson and W. J. Hoefer, *Microwave circuit modeling using electromagnetic field simulation*. Artech House, 2003.

[40] W.-K. Chen, *Theory and design of broadband matching networks*. Pergamon Press, 1976.

[41] ——, *Passive and Active Filters: Theory and Implementations*. John Wiley & Sons, 1986.

[42] H. J. Blinchikoff and A. I. Zverev, *Filtering in the time and frequency domains*. Krieger Publishing Co., 1986.

[43] L. Zhu, S. Sun, and R. Li, *Microwave Bandpass Filters for Wideband Communications*. John Wiley & Sons, 2011.

[44] J.-S. Hong, *Microstrip Filters for RF/Microwave Applications*, 2nd ed. John Wiley & Sons, 2011.

[45] N. Kinayman and M. Aksun, *Modern microwave circuits*. Artech House, 2005.

[46] R. M. Foster, "Academic and theoretical aspects of circuit theory," *Proc. IRE*, vol. 50, no. 5, pp. 866–871, May 1962.

2.21 Exercises

1. The characteristic function of a doubly terminated network is $K(s) = s^2$.
   (a) What is the magnitude-squared transmission coefficient ($|T(s)|^2$)?
   (b) What is the magnitude-squared reflection coefficient ($|\Gamma(s)|^2$)?
   (c) What is transmission coefficient $T(s)$ for a circuit that can be realized using positive R, L, and C elements)?
2. The characteristic function of a doubly terminated network is $K(s) = s^4$. What is the magnitude-squared transmission coefficient ($|\Gamma(s)|^2$)?
3. Consider the design of a fourth-order lowpass Butterworth filter. [This problem follows the development in Section 2.4.]
   (a) What is the magnitude-squared characteristic polynomial, $|K(s)|^2$, of the Butterworth filter?
   (b) What is the magnitude-squared transmission coefficient (or transfer function)?
   (c) What is the magnitude-squared reflection coefficient function?
   (d) Derive the reflection coefficient function (i.e., $\Gamma(s)$). Write down the reflection coefficient in factorized form using up to second-order factors.
   (e) What are the roots of the numerator polynomial of the reflection coefficient function?
   (f) What are the roots of the denominator polynomial of the reflection coefficient function?
   (g) Identify the conjugate pole pairs in the factorized reflection coefficient.
   (h) Plot the poles and zeros of the reflection coefficient on the complex s plane.
4. Derive the reflection coefficient poles of a second-order Butterworth filter and write out the reflection coefficient with nominator and denominator polynomials, that is not in factorized form. [Parallels Example 2.1]
5. Derive the reflection coefficient poles and zeros of a fourth-order Chebyshev filter with a ripple factor, ε, of 0.1.
6. Synthesize the impedance function
$$Z_x = \frac{s^3 + s^2 + 2s + 1}{s^2 + s + 1}.$$
That is, develop the RLC circuit that realizes Z_x. [Parallels Example 2.2]
7. Synthesize the impedance function
$$Z_w = \frac{4s^2 + 2s + 1}{8s^3 + 8s^2 + 2s + 1}.$$
That is, develop the RLC circuit that realizes Z_w. [Parallels Example 2.3]
8. Synthesize the impedance function
$$Z_x = \frac{4s^2 + 2s + 1}{4s^2 + 1}.$$
That is, develop the RLC circuit that realizes Z_x. [Parallels Example 2.2. You may also want to consult Figure 2-9.]
9. Synthesize the impedance function
$$Z_x = \frac{4s^4 + 2s^3 + 5s^2 + 2s + 1}{4s^4 + 4s^3 + 7s^2 + s + 1}.$$
That is, develop the RLC circuit that realizes Z_x. [Parallels Example 2.2. You may also want to consult Figure 2-9.]
10. Develop the lowpass prototype of a fifth-order Butterworth lowpass filter. There may be more than one solution. That is, draw the circuit of the lowpass filter prototype with element values.
11. Develop the lowpass prototype of a fifth-order Chebyshev lowpass filter with 1 dB ripple and 1 rad/s corner frequency.
12. Develop the lowpass prototype of a ninth-order Chebyshev lowpass filter with 0.01 dB ripple and 1 rad/s corner frequency.
13. A 0.04 S admittance inverter is to be implemented in microstrip using a single length of transmission line. The effective permittivity of the line is 9 and the design center frequency is 10 GHz.
   (a) What is the characteristic impedance of the transmission line?
   (b) What is the wavelength in millimeters at the design center frequency in free space?
   (c) What is the wavelength in millimeters at the design center frequency in microstrip?
   (d) What is electrical length of the microstrip transmission line in degrees at the design center frequency?
   (e) What is the length of the microstrip transmission line in millimeters?
14. In Section 2.8.2 it was seen that a series inductor can be replaced by a shunt capacitor with inverters and a negative unity transformer. If the inverter is realized with a one-quarter wavelength long transmission line of characteristic impedance 50 Ω:
   (a) Derive the $ABCD$ parameters of the cascade of Figure 2-16(c) with 50 Ω inverters.

(b) What is the value of the shunt capacitance in the cascade required to realize a 1 nH inductor?

15. A series inductor of 10 pH must be realized by an equivalent circuit using shunt capacitors and sections of one-quarter wavelength long 1 Ω transmission line. Design the equivalent circuit. [Hint: The one-quarter wavelength long lines are impedance inverters.]

16. A series inductor of 10 nH must be realized by an equivalent circuit using shunt capacitors and sections of one-quarter wavelength long 50 Ω transmission line. [Hint: The one-quarter wavelength long lines are impedance inverters.] Design the equivalent circuit.

17. At 5 GHz, a series 5 nH inductor is to be realized using one or more 75 Ω impedance inverters, a unity transformer, and a capacitor. What is the value of the capacitor?

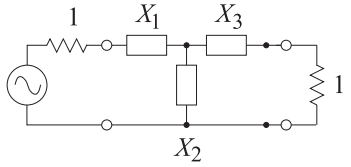
18. In Section 2.8.3 it was seen that a series capacitor can be replaced by a shunt inductor with inverters and a negative unity transformer. Consider that the inverters are realized with a one-quarter wavelength long transmission line of characteristic impedance 100 Ω.

    (a) Derive the $ABCD$ parameters of the cascade of Figure 2-17 with the 100 Ω inverters.
    (b) What is the value of the shunt inductance in the cascade required to realize a 1 pH capacitor?

19. A 50 Ω impedance inverter is to be realized using three resonant stubs. The center frequency of the design is f_0. The first resonant frequency of the stubs is $f_r = 2f_0$.

    (a) Draw the circuit using stubs. On your diagram indicate the input impedance and characteristic impedance of each of the stubs if $f_r = 2f_0$.
    (b) What is the input impedance of a shorted one-eighth wavelength long transmission line if the characteristic impedance of the line is Z_{01}?
    (c) What is the input impedance of an open-ended one-eighth wavelength long transmission
    (d) What is the input impedance of a shorted one-eighth wavelength long transmission if the characteristic impedance of the line is Z_{02}?
    (e) What is the length of each of the stubs in the inverter in terms of the wavelength at the frequency f_0?

20. Design a third-order Type 1 Chebyshev highpass filter with a corner frequency of 1 GHz, a system impedance of 50 Ω, and 0.2 dB ripple. There are a number of steps in the design, and to demonstrate that you understand them you are asked to complete the partial designs indicated below. A Cauer 1 lowpass filter prototype is shown below with ω_c being the corner radian frequency, $f_c = \omega_c/(2\pi)$ being the corner frequency, and Z_0 being the system impedance.



    (a) Design an LPF with $\omega_c = 1$ rad/s, $Z_0 = 1$ Ω.
    (b) Design a HPF with $\omega_c = 1$ rad/s, $Z_0 = 1$ Ω.
    (c) Design a HPF with $f_c = 1$ GHz, $Z_0 = 1$ Ω.
    (d) Design a HPF with $f_c = 1$ GHz, $Z_0 = 50$ Ω.

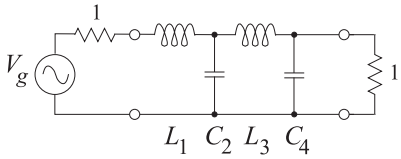
21. This problem considers the design of a Butterworth bandpass filter at 900 MHz.

    (a) Design an LC second-order Butterworth lowpass filter with a corner frequency of 1 rad/s in a 1 Ω system.
    (b) Using the above filter prototype, design a lowpass filter with a corner frequency of 900 MHz.
    (c) Design a second-order Butterworth bandpass filter at 900 MHz using the lowpass filter prototype in (a). Use a fractional bandwidth of 0.1 and a system impedance of 50 Ω.
    (d) What is the 3 dB bandwidth of the filter in (c)?

22. Design a third-order maximally flat bandpass filter prototype in a 50 Ω system centered at 1 GHz with a 10% bandwidth. The lowpass prototype of a third-order maximally flat filter is shown in Figure 2-10.

    (a) Convert the prototype lowpass filter to a lowpass filter with inverters and capacitors only; that is, remove the series inductors.
    (b) Scale the filter to take the corner frequency from 1 rad/s to 1 GHz.
    (c) Transform the lowpass filter into a bandpass filter. That is, replace each shunt capacitor by a parallel LC network. This step will establish the bandwidth of the filter.
    (d) Transform the system impedance of the filter from 1 to 50 Ω.

23. The lowpass prototype of a fourth-order lowpass Butterworth filter is shown below. The corner frequency is 1 rad/s.



$L_1 = 0.765369$ H $\quad C_2 = 1.847759$ F
$L_3 = 1.847759$ H $\quad C_4 = 0.765369$ F

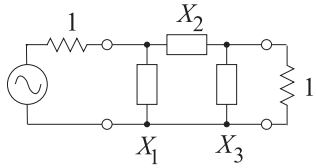
Based on this, develop a fourth-order Butterworth bandpass filter prototype centered at 10^9 rad/s with a fractional bandwidth of 5%.

(a) Scale the lowpass prototype to have a corner frequency of 10^9 rad/s. Draw the prototype with element values.

(b) Draw the schematic of the lumped-element fourth-order Butterworth bandpass prototype based on the original lowpass filter prototype.

(c) Derive the element values of the lumped-element bandpass filter prototype in a 75 Ω system.

24. Design a third-order Type 2 Chebyshev highpass filter with a corner frequency of 1 GHz, a system impedance of 50 Ω, and 0.2 dB ripple. There are a number of steps in the design, and to demonstrate that you understand them you are asked to complete the table below. For each stage of the filter synthesis you must indicate whether the element is an inductance or a capacitance by writing L or C in the appropriate cell. Other cells require a numeric value and you must include units. The X element is identified in the prototype below. A Cauer 2 lowpass filter prototype is shown with ω_c being the corner radian frequency, $f_c = \omega_c/(2\pi)$ being the corner frequency, and Z_0 being the system impedance.



(a) Complete the LPF (lowpass filter) column of the table with $\omega_c = 1$ rad/s, $Z_0 = 1\ \Omega$.

(b) Complete the HPF (highpass filter) column of the table with $\omega_c = 1$ rad/s, $Z_0 = 1\ \Omega$.

(c) Complete the second HPF column of the table with $f_c = 1$ GHz, $Z_0 = 1\ \Omega$.

(d) Complete the third HPF column of the table with $f_c = 1$ GHz, $Z_0 = 50\ \Omega$.

ELEMENT	LPF $\omega_c = 1$ rad/s, $Z_0 = 1\ \Omega$	
	L or C	Value (units)
X_1		
X_2		
X_3		

ELEMENT	HPF $\omega_c = 1$ rad/s, $Z_0 = 1\ \Omega$	
	L or C	Value (units)
X_1		
X_2		
X_3		

ELEMENT	HPF $f_c = 1$ GHz, $Z_0 = 1\ \Omega$	
	L or C	Value (units)
X_1		
X_2		
X_3		

ELEMENT	HPF $f_c = 1$ GHz, $Z_0 = 50\ \Omega$	
	L or C	Value (units)
X_1		
X_2		
X_3		

25. What is the input impedance of a shorted $\lambda/8$ long transmission line with a characteristic impedance of Z_0?

26. What is the input impedance of an open-circuited $\lambda/8$ long transmission line with a characteristic impedance of Z_0?

27. Apply Richards's transformation to a shunt inductor with a reactance of 50 Ω. What is the electrical length of the shorted stub if the stub has a characteristic impedance of 50 Ω?

28. Apply Richards's transformation to a shunt capacitor with a reactance of -50 Ω. What is the characteristic impedance of the open-circuited stub if the electrical length of the stub is one-quarter wavelength long?

2.21.1 Exercises by Section

†challenging, ‡very challenging

§2.2 $1^{\dagger}, 2$
§2.4 $3^{\dagger}, 4^{\dagger}$
§2.5 $5^{\dagger}$
§2.6 $6^{\dagger}, 7^{\dagger}, 8^{\dagger}, 9^{\dagger}$
§2.7 $10, 11^{\dagger}, 12^{\dagger}$
§2.8 $13, 14^{\dagger}, 15^{\dagger}, 16^{\dagger}, 17^{\dagger}, 18^{\dagger}, 19^{\dagger}$
§2.9 $20^{\dagger}, 21^{\dagger}, 2^{\dagger}2, 23^{\dagger}, 24^{\dagger}$
§2.12 $25, 26, 27^{\dagger}, 28^{\dagger}$

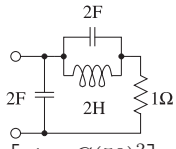
2.21.2 Answers to Selected Exercises

1(c) $\Gamma(s) = \dfrac{s^2}{s^2 + \sqrt{2}s + 1}$

2 $|T(\omega)|^2 = 1/(1 + \omega^8)$

3(f) $-0.38 + \jmath 0.92$
$-0.38 - \jmath 0.92$
$-0.92 + \jmath 0.38$
$-0.38 - \jmath 0.92$

7



14(a) $\begin{bmatrix} 1 & sC(50)^2 \\ 0 & 1 \end{bmatrix}$

13(c) 9
17 889 fF
18 10 nH
19(c) $-Z_2 = -\jmath 50\ \Omega$
24 1 GHz, 50 Ω,
HPF: $X_3 = 6.5$ nH

CHAPTER 3

Parallel Coupled-Line Filters

3.1 Introduction 79
3.2 Coupled Line Configurations 80
3.3 Inverter Network Scaling 91
3.4 Case Study: Third-Order Chebyshev Combline Filter Design 93
3.5 Parallel Coupled Line Filters in an Inhomogeneous Medium ... 114
3.6 Summary 114
3.7 References 115
3.8 Exercises 116

3.1 Introduction

Central to microwave filter design is the identification of a particular distributed structure that inherently has the desired frequency-shaping attributes. Then this structure is adjusted to match an ideal mathematical, or perhaps lumped-element, representation of a filter. Filter design is both science and art, matching synthesis with instinctive knowledge of appropriate physical and circuit structures.

A large number of microwave filters are realized using coupled transmission lines and the most important of these are parallel coupled line (PCL) filters. These are derived from prototypes, with the development following a strategy that enables the filter to be realized using one of several coupled-line configurations. So the strategy is to first examine coupled-line configurations and determine the types of circuit structures that can be realized. Then the steps in synthesis are designed to go from an LC filter prototype to a prototype that has the structures that can be realized by a coupled-line configuration.

Since the beginnings of microwave circuit synthesis, it has become common to present circuit concepts using examples, very often because the steps in realizing a filter may be too difficult to specify algorithmically and there are many steps that require intuition. This procedure is followed here.

This introduction concludes with an example that illustrates the art and science of microwave engineering. It is seen that a simple pair of coupled lines with appropriate matching has a bandpass filter response. The intrinsic bandpass response is supported by coupled lines, so the synthesis procedure is adapting the intrinsic response to specifications.

EXAMPLE 3.1 Coupled Lines as a Basic Element of a Bandpass Filter

AWR Design Environment Project File: RFDesign_Coupled_Shorted_Microstrip_Lines_C.emp

A coupled pair of microstrip lines terminated in short circuits is shown in Figure 3-1. This circuit has a bandpass response centered at 7.5 GHz, as shown in Figure 3-2. So while the insertion loss reduces to about 3.5 dB at 7.5 GHz, the return loss is not very high. This is a good indication that matching could be used to reduce the insertion loss. So the problem is then one of determining the appropriate matching network and to do this an understanding of the circuit response needs to be gained. Figure 3-3 plots the S_{11} and S_{21} responses on a Smith chart. By looking at S_{11} it is seen that between 7 and 8 GHz the input impedance is primarily inductive. This indicates that a series capacitance could be used in matching. Another view is shown in Figure 3-4, which plots the magnitude and phase of the input impedance at Port 1 of the coupled microstrip lines (this will be the same as the input impedance at Port 2). So around 7 GHz, the input impedance is inductive since the phase of the impedance is close to 90° and the magnitude of the impedance is increasing with frequency. At 7 GHz the impedance is approximately $j50\ \Omega$, which would be resonated out by a series capacitor of 0.45 pF. However it will not be as simple as this, as the tuning capacitor will need to be placed at both Ports 1 and 2. However, this does serve to provide an initial design point.

The schematic required to implement the design above is shown in Figure 3-5. In this circuit schematic, the coupled line shown in Figure 3-1(b) is captured as a subcircuit called "CoupledLine." The series capacitors at Ports 1 and 2 are "C1" and "C2," that is, C_1 and C_2, respectively. C_1 and C_2 are established as tunable elements with the governing variable being "CC." This capacitor is tuned to obtain the optimum bandpass response. With $C_1 = C_2 = 44.8$ fF, the bandpass response shown in Figure 3-6 is obtained. This is an almost ideal maximally flat bandpass filter response. The main passband is at 7 GHz and there is a parasitic bandpass response at the third harmonic. This spurious passband at an odd harmonic occurs often in transmission line designs. A small alteration of the tuning capacitor can change the response to have ripples in the passband and sharper filter skirts (see Figure 3-7), where the tuning capacitors are each 37.6 fF. (To convince yourself of the sharper skirt consider the insertion loss at 100 MHz away from the passband. The insertion loss of the filter with the maximally flat response is 23 dB there, and that of the filter with the ripple response is 27 dB.) This filter has two passband poles and is simple enough to design as done here. Higher-order filters (with more than two passband resonators) require a more sophisticated design approach. Still this example demonstrates that the coupled-line structure has a good passband response on its own provided that appropriate matching is used.

The bandpass characteristic is a result of the phase velocities of the even and odd modes being different. If they were the same, the transmission coefficient would be zero at 7 GHz.

3.2 Parallel Coupled Lines

Several of the key advantages of coupled-line filters over other distributed filters is that the coupling between resonators can be strong if the pair of coupled-line resonators are close to each other, or low if they are widely separated. This enables filters to be realized with high (30–70%), medium (10–30%), or low (2–10%) bandwidths. The low-bandwidth filters require low coupling of the resonators, while the high bandwidth filters require high coupling. Coupled-line filters are also compact and can be implemented in microstrip configurations or most of the other transmission line technologies. The different transmission line forms have different Qs. Planar transmission lines have moderate Qs especially at high frequencies, due principally to

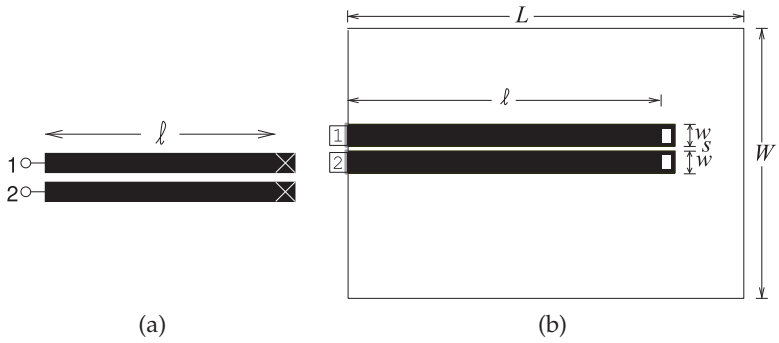


Figure 3-1: Coupled microstrip line layout: (a) schematic; and (b) layout in an EM simulator. Dimensions of the coupled lines are w = 500 μm, s = 100 μm, ℓ = 1 cm, W = 6 mm, and L = 12 mm. The metal is 6 μm thick gold (conductivity $\sigma = 42.6 \times 10^6$ S/m) and the alumina substrate height is 600 μm with relative permittivity ε_r = 9.8 and loss tangent of 0.001.

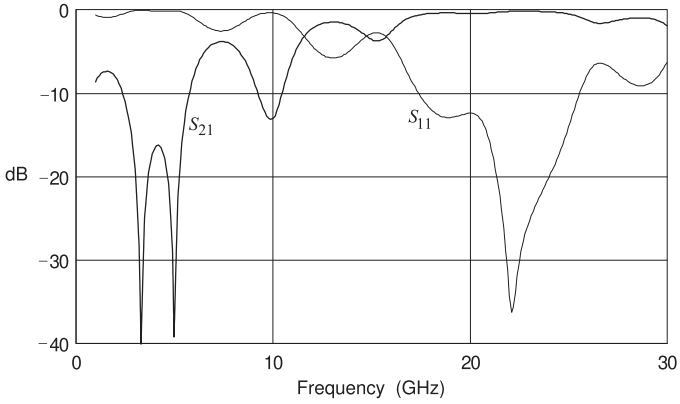


Figure 3-2: Insertion loss (S_{21}) and return loss (S_{11}) of the coupled line of Figure 3-1 with an $s = 100$ μm gap.

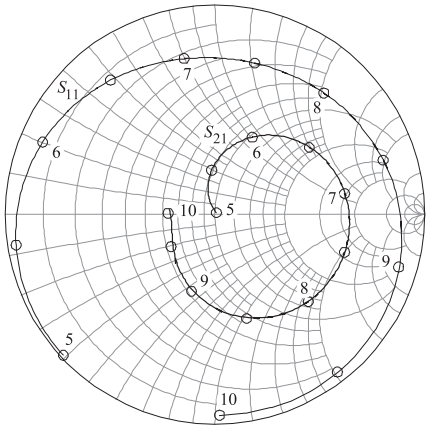


Figure 3-3: Insertion loss and return loss over a narrow frequency range plotted on a Smith chart. The annotation at the markers is the frequency in GHz.

radiated fields. Low Q means that the insertion loss of filters will be high and the filter skirts will not be as steep as they would be if a high-Q transmission line structure was used. For example, up to the early 2000s the RF filters in the front end of cellular phones were mostly seventh-order Chebyshev filters using coupled slablines as shown in Figure 3-8 (although here only

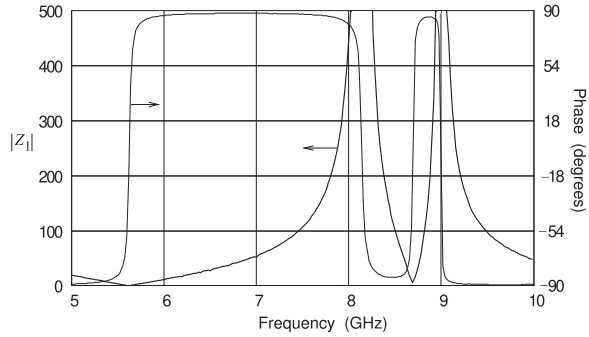
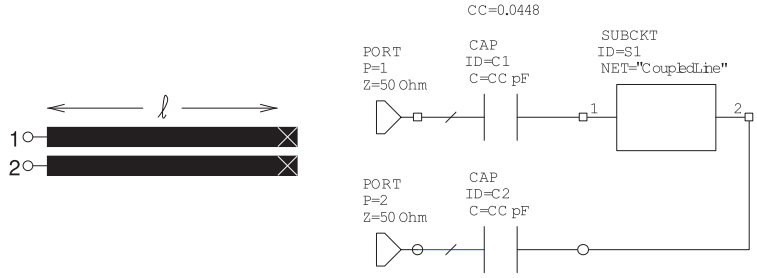
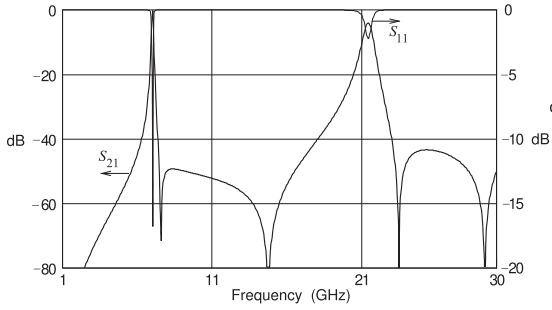


Figure 3-4: Input impedance at Port 1 of the coupled microstrip lines in Figure 3-1. Port 2 is terminated in 50 Ω.

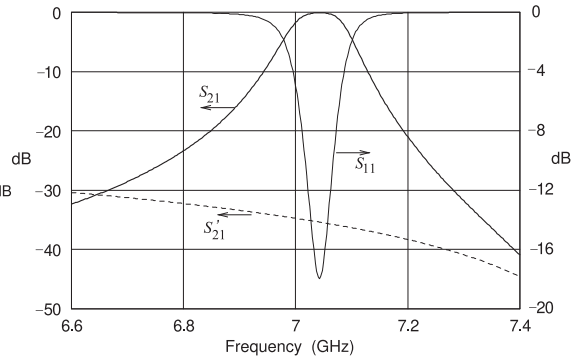


(a) (b)

Figure 3-5: Capacitively coupled microstrip lines: (a) schematic; and (b) representation in a microwave CAD tool with a subcircuit representing the coupled lines. The layout of the subcircuit is shown in Figure 3-1.



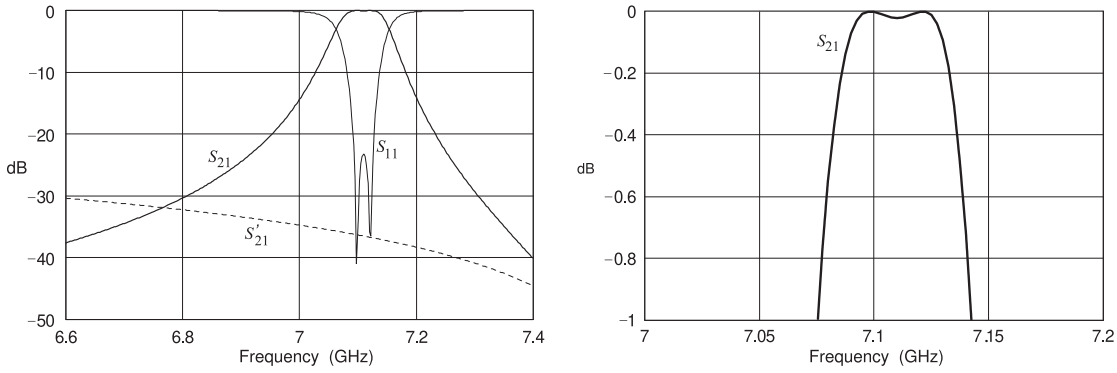
(a) Wide frequency view



(b) Detail

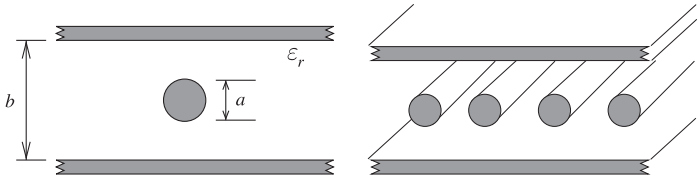
Figure 3-6: Return loss and insertion loss when the coupling capacitors are 44.8 fF. This results in a Butterworth-like response. S'_{21} is the response without C_1 and C_2.

four resonators are shown for clarity). Typically the top and bottom plates of the resonators were 2–3 mm apart and the area of the slabline filter was 1 cm×1 cm. This is too large for today's thin smart phones. However, they are very good filters and cheap to produce. The electrical design procedure for parallel coupled slabline filters and for using other transmission line structures is the same as for parallel coupled microstrip filters. The difference is only in the final physical implementation.



(a) Narrowband response (b) Detail showing ripples

Figure 3-7: Return loss and insertion loss when the coupling capacitors are 37.6 fF. This results in a Chebyshev-like response. S'_{21} is the response without C_1 and C_2.



(a) (b)

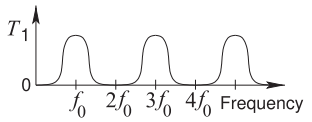
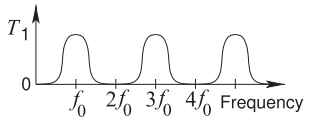
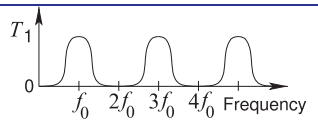
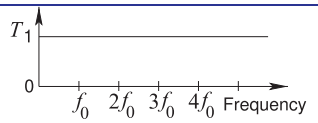
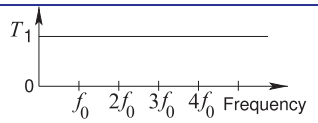
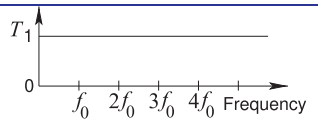
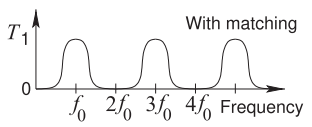
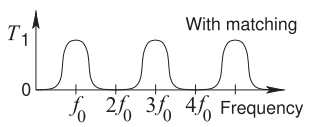
Figure 3-8: Slabline: (a) cross section of a single slabline; and (b) parallel coupled-line configuration with four coupled slablines.

3.2.1 Coupled-Line Configurations

The coupled-line configuration considered in Example 3.1 is called a combline section. This is one of many that have desirable frequency-selective responses. The characteristics of many parallel coupled-line configurations having bandpass, all-pass, or all-stop characteristics are shown in Table 3-1. The most important configurations are starred ($\star$).

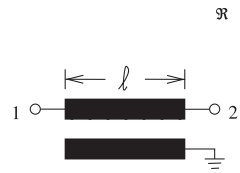
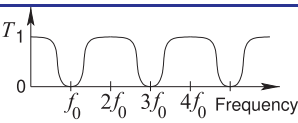
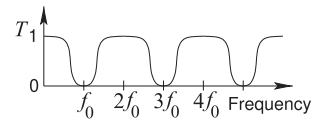
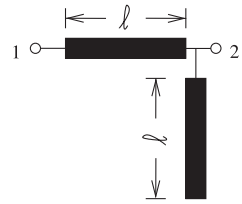
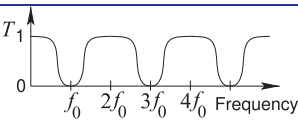
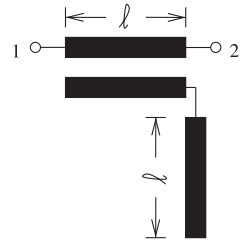
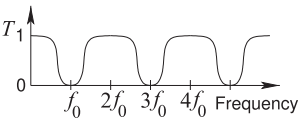
The PCL section of Filter (a) in Table 3-1 is called an interdigital section and is a bandpass filter. The dual filter is Filter (b), with what is called a parallel coupled section, and both can be used to realize narrow (2–3%) to wide (up to 30%) bandwidth filters with close to symmetrical responses around the center frequency f_0. For a wide bandwidth filter the lines are close and for a narrow bandwidth filter the parallel lines are widely separated so that there is low-level coupling. Filter (c) is also a bandpass filter, however, the frequency selectivity is not as good as with Filters (a) and (b). Filters (d), (e), and (f) are all-pass filters. These filters are used to adjust the phase of a transmitted signal, usually to accommodate phase dispersion (e.g., phase variation with respect to frequency) that was introduced somewhere else. While this analog function was once important, it is less so now as DSP techniques can usually equalize the phase sufficiently. Filter (g) is an all-stop filter, meaning that negligible signal is transmitted. This filter has no practical use. Filters (h) and (i) are inherently all-stop filters. An example of the response of the combline section of Filter (h) was seen in Figure 3-2 of Example 3.1, where the all-stop response is centered at 4 GHz. The response is too broad to be useful as a bandstop filter (in any practical application).

Table 3-1: Responses of the nine coupled-line configurations with narrow (2–3%), moderate (3–10%), and wide (10%–30%) bandwidth. The most important configurations are starred (⋆).

Attributes	Circuit	Response
⋆ (a) Interdigital section. Bandpass. Narrow to wide bandwidth. $\ell = \lambda/4$ at f_0. Symmetrical passband.		
⋆ (b) Parallel coupled section. Bandpass. Narrow to wide bandwidth. $\ell = \lambda/4$ at f_0. Symmetrical passband.		
(c) Bandpass. Moderate to wide bandwidth. $\ell = \lambda/4$ at f_0.		
(d) All-pass. $0 < \ell \leq \lambda/2$ at f_0.		
(e) All-pass. $0 < \ell \leq \lambda/2$ at f_0.		
(f) All-pass. $0 < \ell \leq \lambda/2$ at f_0.		
(g) All-stop (no practical use).		
⋆ (h) Combline section. All-stop without matching. Bandpass with matching. Moderate to wide bandwidth. $\ell = \lambda/4$ at f_0. Compact. Asymmetrical response.		
(i) All-stop without matching. Bandpass with matching. Moderate to wide bandwidth. $\ell = \lambda/4$. Compact. Asymmetrical response.		

However, with minimal matching, such as a series capacitor at each of ports (see Figure 3-5), the combline section becomes a very good bandpass filter (see Figures 3-6 and 3-7). The combline bandpass filter, Filter (h), is more compact than Filters (a) and (b). However, it has an asymmetrical bandpass response (which can be seen in Figure 3-7(a)), whereas Filters (a) and (b) have a symmetrical frequency response. Which type of response is preferred depends on the application. Filter (i) is the dual of Filter (h), but in practice

Table 3-2: Responses of coupled-line configurations having lowpass and bandstop responses.

Attributes	Circuit	Response
⋆ (j) Lowpass. Narrow to moderate bandwidth. $\ell = \lambda/2$ at f_0		
⋆ (k) Lowpass. Narrow to moderate bandwidth. $\ell = \lambda/4$ at f_0		
⋆ (l) Lowpass. Narrow bandwidth. $\ell = \lambda/4$ at f_0.		
⋆ (m) Bandstop. Narrow bandwidth. $\ell = \lambda/4$ at f_0.		

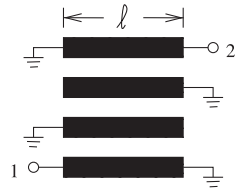
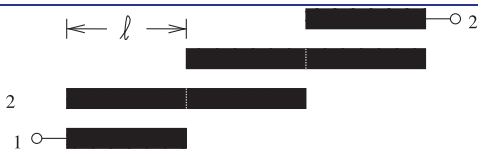
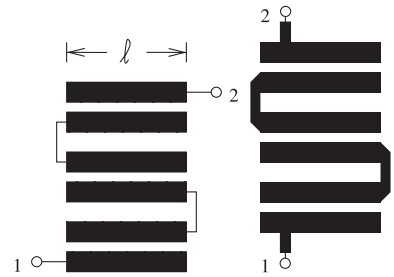
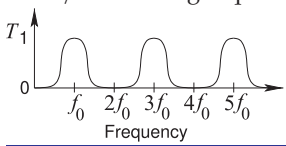
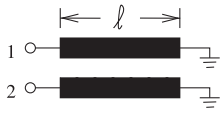
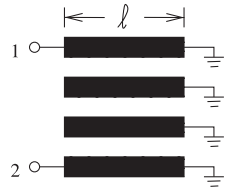
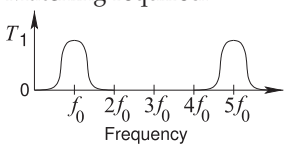
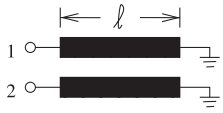
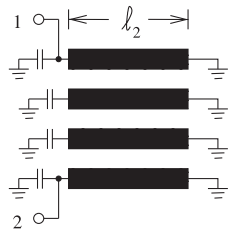
the combline configuration of Filter (h) is preferred.

Table 3-2 presents coupled-line sections having lowpass and bandstop responses. All of these filters have desirable characteristics and so all are starred (⋆). Filter (l) is not a PCL filter but it is used in trade-off among the three lowpass filters (i.e., Filters (j), (l) and (m)). The usual application of a bandstop filter is to notch out an undesired signal (e.g. to prevent an LO from appearing where it is not desired). As such, the desired frequency response is narrowband and Filter (m) is a good choice for a bandstop filter.

Filter design using PCL sections begins with the choice of a PCL configuration having the essential desired frequency response. Each of the PCL sections in Tables 3-1 and 3-2 have two resonators and can implement bandpass, bandstop, and lowpass filters having second-order responses.[1] To meet specifications it is generally necessary to replicate the basic section. So treating the two resonator configurations as unit cells, a multi-cell filter can be realized. The multi-cell forms of the main bandpass configurations are shown in Table 3-3. The multi-cell implementation of the interdigital, parallel

[1] Recall that the order designation comes from the lowpass prototype so that here a second-order bandpass filter actually has two resonators, each having an LC-like response.

Table 3-3: Multi-cell forms of bandpass parallel coupled-line configurations. The wavelength, λ, is at the center frequency of the filter.

Name	Unit cell	Multi-cell form
(a) Interdigital bandpass filter. $\ell = \lambda/4$.		
(b) Parallel edge-coupled bandpass filter. $\ell = \lambda/4$.		
(c) Parallel edge-coupled hairpin bandpass filter. $\ell > \lambda/4$.		
(d) Combline bandpass filter. $\ell = \lambda/4$. Matching required. 		
(e) Combline bandpass filter with extended stopband. $\ell = \lambda/4$. $\ell_2 \approx \lambda/8$ (typically). Matching required. 		

edge-coupled bandpass filters, (a) and (b) in Tables 3-3, is straightforward. A variation is shown in Filter (c), which folds the parallel edge-coupled sections to realize a compact multi-cell form called a hairpin filter [1–3]. One of the consequences of using transmission line segments to realize a filter is that there are spurious passbands. This results because a $\lambda/4$ section of line looks the same electrically as a $3\lambda/4$ section. Consequently there are normally spurious passbands at the odd-harmonic frequencies. If the basic resonator is $\lambda/2$ long, then the spurious passbands would be at all of the harmonic frequencies. A solution is to incorporate capacitors in the

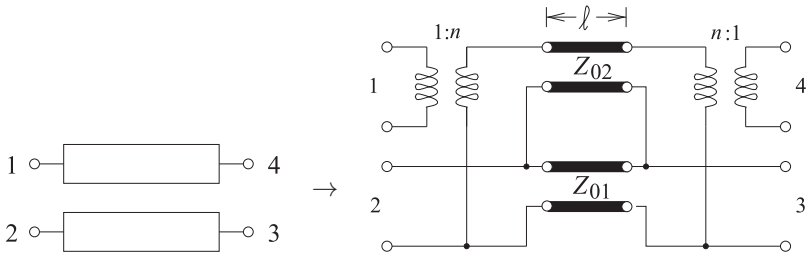


Figure 3-9: Network model of a pair of coupled lines.

resonators as shown in Filter (e) in Table 3-3 [2, 4, 5]. Thus each of the original transmission line resonators now becomes a combination of a capacitor and a shorter transmission line segment. If the new transmission line segment is $\lambda/8$ long, the first spurious passband will now be at $5f_0$ rather than $3f_0$. The spurious passbands can be pushed further up in frequency by using even shorter transmission line lengths, but the performance of the filter at the passband frequency, f_0, will be compromised.

There are many variations on the PCL filter and several different design techniques have been developed [6–28]. Refer to these citations to explore alternative PCL configurations and alternative methods of design to that presented here. This chapter presents one of the common approaches to synthesizing PCL filters and the scheme accommodates the use of capacitive loading used to extend the stopband of a passband filter.

Before launching into the synthesis procedure for a PCL filter, a comment on synthesis versus optimization is warranted. With a simple topology such as the second-order filter topologies shown in Table 3-1, an optimization procedure could be used to design the widths and lengths of the lines and the two or four variables describing the required external matching networks that are not shown. A global optimization is certainly feasible. However, it is not feasible to use global optimization of, for example, a seventh-order filter required to meet typical cell phone specifications. An exception is if a design for a very similar specification is available and the changes required are small. Competitive filter design requires synthesis. This will lead to optimum performance, and the insight gained can be used in topology modifications.

3.2.2 Coupled-Line Circuit Models

An approximate coupled-line model of a pair of coupled lines was presented in Section 5.9.5 of [29]. This model is repeated in Figure 3-9. The parameters of the network model are related to the model impedances as follows:

$$n = \frac{1}{K} = \frac{Z_{0e} + Z_{0o}}{Z_{0e} - Z_{0o}} \qquad (3.1)$$

$$Z_{0S} = \sqrt{(Z_{0e} Z_{0o})}, \qquad (3.2)$$

$$Z_{01} = \frac{Z_{0S}}{\sqrt{1 - K^2}}, \qquad (3.3)$$

$$Z_{02} = Z_{0S} \frac{\sqrt{1 - K^2}}{K^2}. \qquad (3.4)$$

Various terminating arrangements of the coupled lines result in several useful filter elements. One arrangement is shown in Figure 3-10. Also shown in this figure is the development of the network model based on the model in Figure 3-9. The final network model is a transmission line of characteristic impedance Z_{01} in cascade with an open-circuited stub. Consider what happens at the resonant frequency, f_r (the frequency at which the lines are one-quarter wavelength long). At lower frequencies, $f \ll f_r$, the Z_{02} line

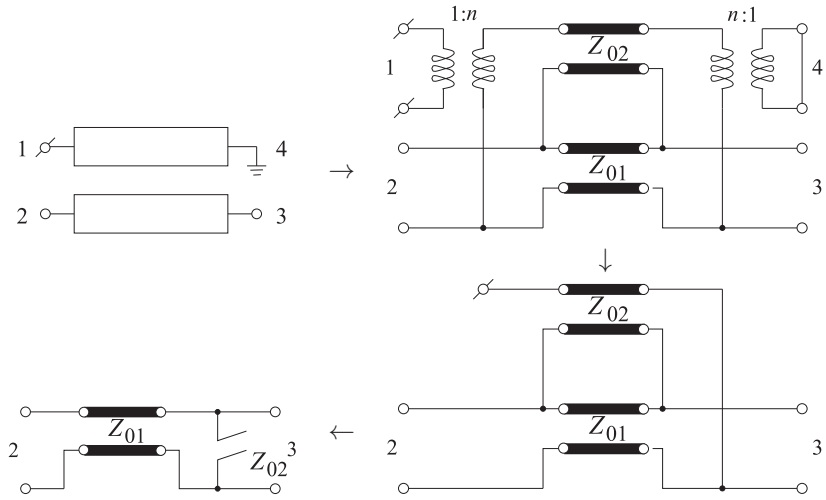


Figure 3-10: Lowpass distributed network section derived from a pair of coupled lines with Port 1 open-circuited. The open circuit is indicated by a node (open circle) with a line through it. The final network model is a transmission line of characteristic impedance Z_{01} and an open-circuited stub of characteristic impedance Z_{02}. The lines and stubs are one-quarter wavelength long at the corner frequency. (Thus with the stub $f_r = f_0$, and the characteristic impedance of the stub is as shown.)

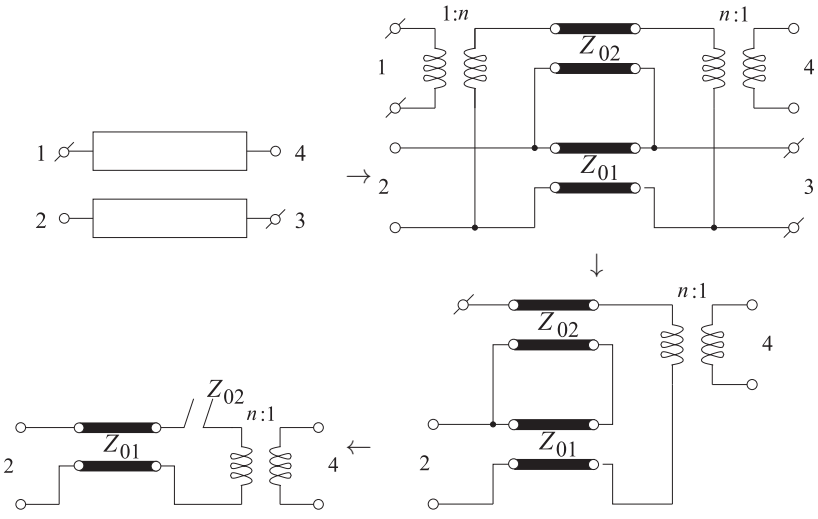


Figure 3-11: Parallel coupled-line section with Ports 1 and 3 open-circuited and network models. (For the stub, the characteristic impedance of the stub is shown and $f_r = f_0$.)

is an open circuit and signals travel along Z_{01}. At resonance the Z_{02} stub becomes a short circuit and signals do not pass. This is a crude verification that this is a lowpass structure. The process is visual and is expected to be self-explanatory. Other examples are shown in Figures 3-11 to 3-13.

The model of a combline section is shown in Figure 3-13. The final network reduction is repeated in Figure 3-14, and it will be shown that the model in Figure 3-14(b) is equivalent to the model in Figure 3-14(a). In the synthesis of a combline filter, the network of Figure 3-14(b) is obtained and this can be

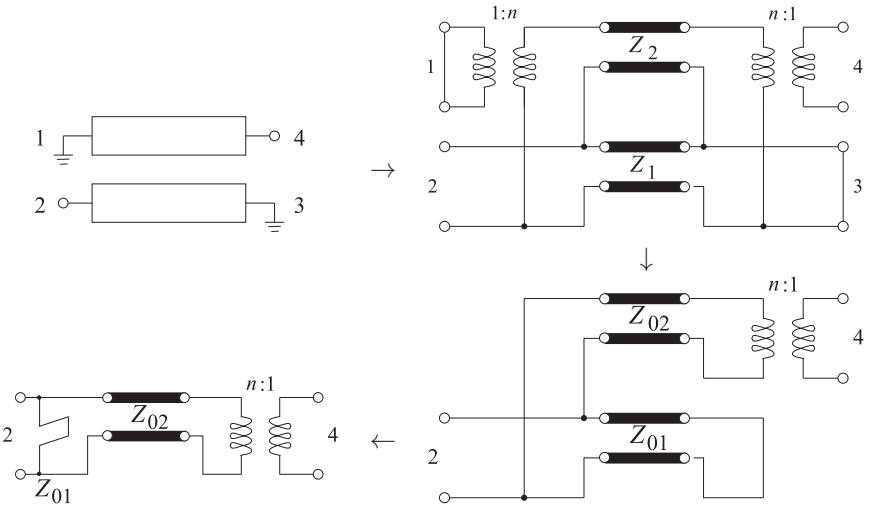


Figure 3-12: Interdigital section and network models. (For the stub, the characteristic impedance of the stub is shown and $f_r = f_0$.)

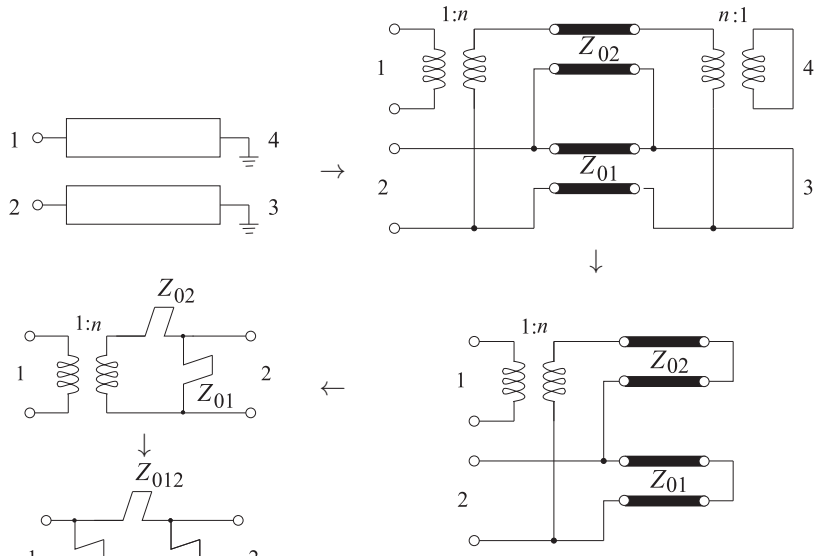


Figure 3-13: Combline section and network models. (For the stubs, the characteristic impedances of the stubs are shown and $f_r = f_0$.)

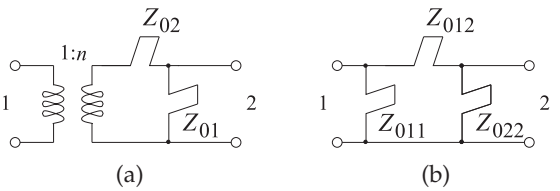


Figure 3-14: Equivalent models of a section of combline. (For the stubs, the characteristic impedances of the stubs are shown and $f_r = f_0$.)

related back to the dimensions of the coupled line. The equivalence is done using $ABCD$ parameters and, as will be seen, the equivalence will not be at just one frequency but will be broadband. The $ABCD$ parameters of the network in Figure 3-14(a) are obtained by cascading the $ABCD$ parameters of three two-ports (the $ABCD$ parameters of which are given in Table 2-1 of [30]). The $ABCD$ parameters of the network in Figure 3-14(a) are (from

the multiplication of three $ABCD$ parameter matrices)

$$T_A = T_{\text{TRANSFORMER}} T_{\text{SERIES STUB}} T_{\text{SHUNT STUB}} \tag{3.5}$$

$$= \begin{bmatrix} 1/n & 0 \\ 0 & n \end{bmatrix} \begin{bmatrix} 1 & \jmath Z_{02} \tan\theta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\jmath/(Z_{01}\tan\theta) & 1 \end{bmatrix} \tag{3.6}$$

$$= \begin{bmatrix} 1/n & \jmath Z_{02} \tan\theta/n \\ 0 & n \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\jmath/(Z_{01}\tan\theta) & 1 \end{bmatrix} \tag{3.7}$$

$$= \begin{bmatrix} \dfrac{1}{n}\left(1+\dfrac{Z_{02}}{Z_{01}}\right) & \jmath Z_{02}\tan\theta/n \\ -\jmath n/(Z_{01}\tan\theta) & n \end{bmatrix}. \tag{3.8}$$

The $ABCD$ parameters of the network in Figure 3-14(b) are

$$T_B = T_{\text{SHUNT STUB}} T_{\text{SERIES STUB}} T_{\text{SHUNT STUB}} \tag{3.9}$$

$$= \begin{bmatrix} 1 & 0 \\ -\jmath/(Z_{011}\tan\theta) & 1 \end{bmatrix} \begin{bmatrix} 1 & \jmath Z_{012} \tan\theta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\jmath/(Z_{022}\tan\theta) & 1 \end{bmatrix} \tag{3.10}$$

$$= \begin{bmatrix} 1 & \jmath Z_{012}\tan\theta \\ -\jmath/(Z_{011}\tan\theta) & 1+Z_{012}/Z_{011} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\jmath/(Z_{022}\tan\theta) & 1 \end{bmatrix} \tag{3.11}$$

$$= \begin{bmatrix} 1+\frac{Z_{012}}{Z_{022}} & \jmath Z_{012}\tan\theta \\ \dfrac{-\jmath}{\tan\theta}\left(\dfrac{1}{Z_{011}}+\dfrac{1}{Z_{022}}+\dfrac{Z_{012}}{Z_{011}Z_{022}}\right) & 1+\dfrac{Z_{012}}{Z_{011}} \end{bmatrix}. \tag{3.12}$$

Equating Equations (3.8) and (3.12) yields

$$1+\frac{Z_{012}}{Z_{022}} = \frac{1}{n}\left(1+\frac{Z_{02}}{Z_{01}}\right) \tag{3.13}$$

$$Z_{012} = Z_{02}/n \tag{3.14}$$

$$\left(\frac{Z_{011}+Z_{022}+Z_{012}}{Z_{011}Z_{022}}\right) = \frac{n}{Z_{01}} \tag{3.15}$$

$$1+\frac{Z_{012}}{Z_{011}} = n, \tag{3.16}$$

which have the solution

$$Z_{012} = \frac{Z_{02}}{n} \tag{3.18}$$

$$Z_{011} = \frac{Z_{012}}{n-1} = \frac{Z_{02}}{n(n-1)} \tag{3.17}$$

$$Z_{022} = \frac{Z_{01}Z_{02}}{Z_{02}-(n-1)Z_{01}}. \tag{3.19}$$

Rearranging these, expressions for Z_{01}, Z_{02}, and n can be obtained:

$$n = 1+\frac{Z_{012}}{Z_{011}},\ Z_{01} = \left(\frac{nZ_{011}Z_{022}}{Z_{011}+Z_{022}+Z_{012}}\right), \text{ and } Z_{02} = nZ_{02}Z_{012}. \tag{3.20}$$

Using these, Equations (3.1)–(3.4), and the coupled-line analysis of Section 5.6 of [29], the geometric parameters of the combline coupled-line section can be obtained corresponding to the stub circuit of Figure 3-14(b).

Thus the equivalent circuit of the combline section, the top left figure in Figure 3-13, has the equivalent circuit shown in Figure 3-14. Filter synthesis can be directed at developing circuit structures like that in Figure 3-14(b) and from this electrical design, the physical design consisting of combline sections can be developed.

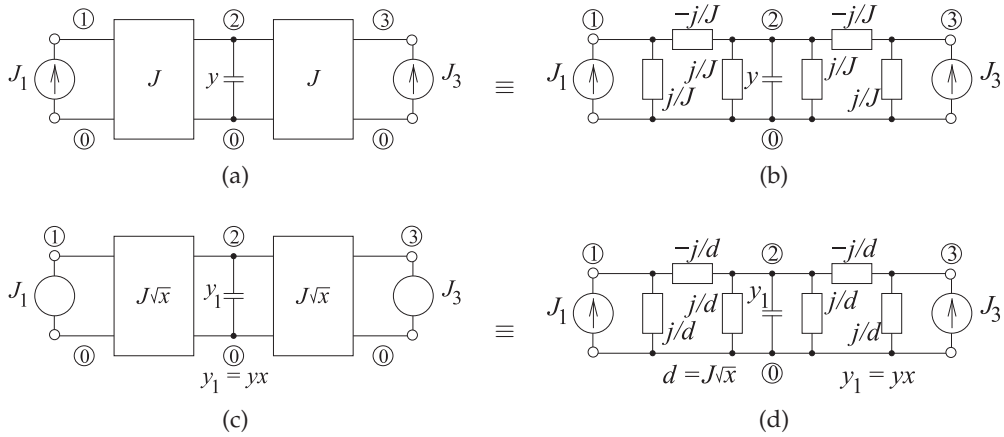


Figure 3-15: Inverter network: (a) network with two identical admittance inverters with an inserted shunt element of admittance, y; (b) equivalent network using equivalence shown in Figure 2-19; (c) scaled original network; and (d) scaled equivalent network. Element values are impedances except for J, y and y_1, which are admittances. The use of the equivalent networks in filter synthesis is illustrated in the case study in Section 3.4.

3.3 Inverter Network Scaling

A filter that uses transmission lines is nearly always synthesized from a filter prototype that contains inverters. The synthesis of a filter begins with a normalized prototype that is transformed to the desired frequency, impedance, and type. It is the purpose of this section to show how to use such scaling when there are inverters in the prototype. In particular, it is shown that scaling the admittance of the network of Figure 3-15(a), a common inverter subnetwork, by a factor x results in the network in Figure 3-15(c).

The nodal admittance matrix of the network in Figure 3-15(a) is

$$Y = \begin{bmatrix} 0 & -\jmath J & 0 \\ -\jmath J & y & -\jmath J \\ 0 & -\jmath J & 0 \end{bmatrix}. \tag{3.21}$$

Assigning nodal voltages to Terminals 1, 2, and 3, the nodal admittance matrix equation is

$$\begin{bmatrix} 0 & -\jmath J & 0 \\ -\jmath J & y & -\jmath J \\ 0 & -\jmath J & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} J_1 \\ 0 \\ J_3 \end{bmatrix}, \tag{3.22}$$

and by eliminating Node 2 using network condensation (see Section 1.A.16) of [31], this reduces to

$$\begin{bmatrix} J^2/y & J^2/y \\ J^2/y & J^2/y \end{bmatrix} \begin{bmatrix} v_1 \\ v_3 \end{bmatrix} = \begin{bmatrix} J_1 \\ J_3 \end{bmatrix}, \tag{3.23}$$

and this describes the external characteristics of the subnetwork in Figure 3-15(a).

The nodal admittance matrix of the scaled network in Figure 3-15(c) is

$$Y' = \begin{bmatrix} 0 & -\jmath J\sqrt{x} & 0 \\ -\jmath J\sqrt{x} & yx & -\jmath J\sqrt{x} \\ 0 & -\jmath J\sqrt{x} & 0 \end{bmatrix}. \tag{3.24}$$

That is,

$$\begin{bmatrix} 0 & -\jmath J\sqrt{x} & 0 \\ -\jmath J\sqrt{x} & yx & -\jmath J\sqrt{x} \\ 0 & -\jmath J\sqrt{x} & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} J_1 \\ 0 \\ J_3 \end{bmatrix}, \tag{3.25}$$

and by eliminating Node 2 this reduces to

$$\begin{bmatrix} (J^2x)/(yx) & (J^2x)/(yx) \\ (J^2x)/(yx) & (J^2x)/(yx) \end{bmatrix} \begin{bmatrix} v_1 \\ v_3 \end{bmatrix} = \begin{bmatrix} J^2/y & J^2/y \\ J^2/y & J^2/y \end{bmatrix} \begin{bmatrix} v_1 \\ v_3 \end{bmatrix} = \begin{bmatrix} J_1 \\ J_3 \end{bmatrix}. \tag{3.26}$$

Thus the original network shown in Figure 3-15(a) has the same external electrical characteristics as the scaled network of Figure 3-15(c), with the characteristic admittance of the inverters scaled by $\sqrt{x}$ and the shunt admittance scaled by x.

A generalization of this result (which is useful when there are additional connections between Nodes 1 and 3) is that multiplying a row and a column of the nodal admittance matrix by the same factor results in identical external characteristics. Note that the element sharing a row and column is multiplied twice.

EXAMPLE 3.2 Inductor Synthesis

Lumped inductors have low Q. Fortunately, in microwave design they can be realized using inverters and a shunt capacitor that have a high Q. Realize the series inductor in Figure 3-16(a) with a shunt capacitor and 10 Ω impedance inverters.

Solution:

The series inductor in a 1 Ω system is transformed into a network with inverters and a shunt capacitor using the transformation shown in Figure 2-48. Thus here the series inductor can be realized by the circuit of Figure 3-16(b), where the shunt capacitor C has the same numeric value as the series inductor. The inverters here can be either impedance inverters or admittance inverters, as their value is equal to one. The design requires that these be realized as 10 Ω inverters, but scaling is performed on admittance inverters, as shown in Figure 3-16(c), where the value of the admittance inverters is 0.1 S = $\sqrt{x}$. Thus $x = 0.01$ and the admittance of the capacitor is scaled by 0.01, so $C_1 = C/100$ = 10 pF. The final design is shown in Figure 3-16(d).

3.4 Case Study: Third-Order Chebyshev Combline Filter Design

This section presents the design of a 1 GHz third-order Chebyshev combline bandpass filter with 10% bandwidth. The combline filter is one of the most commonly used transmission line-based filters and uses the combline section shown in Figure 3-13. The synthesis that will be presented here combines many of the concepts discussed previously in this chapter. There are also design decisions that must be made in response to particular situations encountered during synthesis. An example of a design decision is deciding what to do if the characteristic impedance of a transmission line is too low.

Step 1: Choice of Lowpass Filter Prototype

Design begins with the choice of a prototype that is expected to achieve the required specifications. If the choice is eventually found to be not suitable, then another topology must be chosen and the synthesis restarted. Consider a third-order Chebyshev lowpass filter prototype, shown in Figure 3-17, with a passband ripple factor of $\varepsilon = 0.1$ (or 0.043 dB). The element values were calculated using Equations (2.83)–(2.90). Note that at the corner frequency that defines the bandwidth of the lowpass filter, the square of the transmission coefficient is the ripple level (see Figure 2-7). That is, at frequency $\omega_0 = 1$ rad/s, the squared transmission coefficient is $1/(1+\varepsilon^2)$ (or in decibels the ripple is $R_{\text{dB}} = 10\log(1+\varepsilon^2)$).

Step 2: Replace Series Elements

The next step in the transformation of the lowpass prototype in Figure 3-17 is replacing the series inductor by a shunt capacitor in cascade with admittance inverters, as shown in Figure 3-18. This step uses the series inductor transformation shown in Figure 2-16.

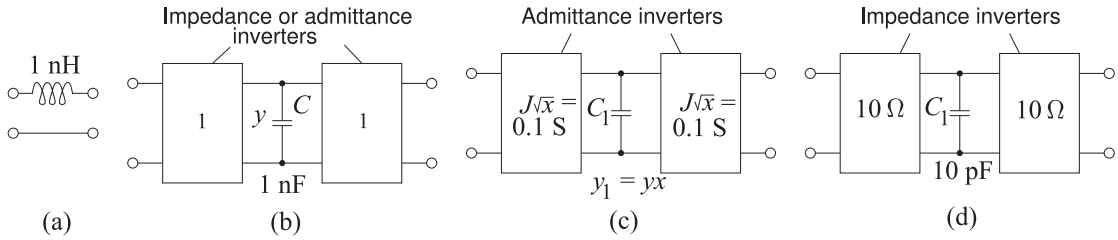
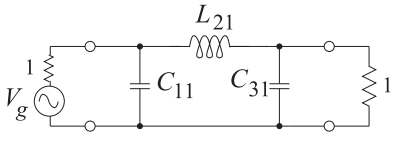
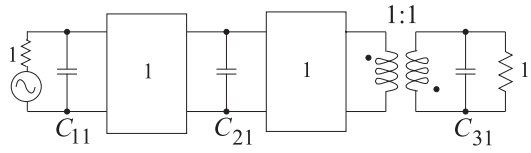


Figure 3-16: Realization of a series inductor as a shunt capacitor with inverters.

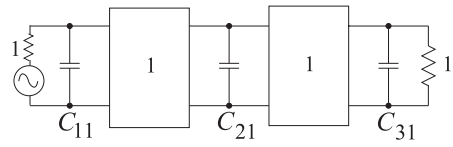


$C_{11} = 0.85158$ F $= C_{31}$ $L_{21} = 1.10316$ H

Figure 3-17: Step 1. A third-order Chebyshev lowpass filter prototype.
$g_1 = g_3 = 0.85158$, $g_2 = 1.10316$.



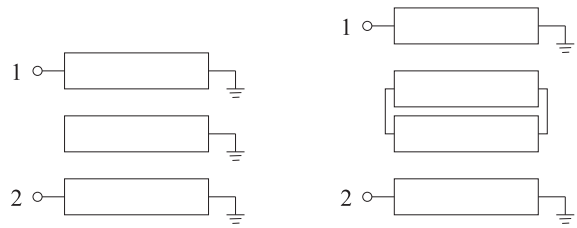
(a)



(b)

$C_{11} = 0.85158$ F $= C_{31}$ $C_{21} = 1.10316$ F

Figure 3-18: Step 2. Prototype filter with inductor replaced by a capacitor and admittance inverter combination: (a) with a unity inverting transformer; and (b) with the transformer eliminated as it shifts the transmission phase by 180° and so has no effect on the filter response. Note that $|C_{21}| = |L_{21}|$.



(a) Three-resonator combline filter (b) Model

Figure 3-19: Target combline filter physical layout.

Step 3: Bandpass Filter Transformation

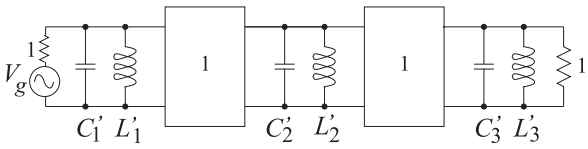
In synthesizing the bandpass filter the final physical topology must be considered. The essential element of a third-order combline filter comprises three coupled lines as shown in Figure 3-19(a). The three resonators can be modeled as two pairs of coupled lines, as shown in Figure 3-19(b), with one line shared by each pair. Each pair of coupled lines can be modeled by a Pi network of stubs as shown in Figure 3-14(b). Thus the lowpass prototype of Figure 3-18(b) needs to be converted to a bandpass filter and eventually this needs to be transformed into two cascaded Pi networks of stubs.

In this step the prototype is converted to a bandpass circuit. The filter has 10% bandwidth with approximate corner frequencies at $f_1 = 950$ MHz and $f_2 = 1050$ MHz. Using the filter type transformations in Figure 2-27, $\omega_0 = 2\pi 10^9$, and $\alpha = \omega_0/(\omega_2 - \omega_1)$ = 1/(fractional bandwidth) = 10, the first capacitor in Figure 3-18, C_{11}, is transformed to

$$C_1' = \frac{\alpha C_{11}}{\omega_0} = 1355.33 \text{ pF} \text{ in parallel with } L_1' = \frac{1}{\alpha C_{11}\omega_0} = 0.0186894 \text{ nH}. \tag{3.27}$$

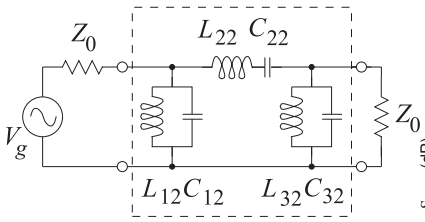
The other capacitors are transformed similarly and the bandpass prototype becomes that shown in Figure 3-20.

The lumped-element version of the bandpass filter is shown in Figure 3-21 along with its simulated response. This response is the ideal electrical response and provides a reference in design. Note that the ripples in the transmission response, i.e. S_{21}, are not evident on this scale. Losses in the implemented filter will smooth the ripples even further although the poles in the S_{21} response are not evident, these correspond to the zeros in the S_{11} response and these are clearly visible. Figure 3-21 shows three zeros in the S_{11} response, as expected for a third-order filter.



$C_1' = 1355.33\,\text{pF} = C_3' \quad L_1' = 0.0186894\,\text{nH} = L_3'$
$C_2' = 1755.73\,\text{pF} \quad L_2' = 0.0144271\,\text{nH}$

Figure 3-20: Step 3. Bandpass filter prototype with inverters derived from the lowpass filter prototype of Figure 3-18.



$Z_0 = 1\ \Omega$
$C_{12} = 1.35533\,\text{nF} = C_{32}$
$L_{12} = 18.6894\,\text{pH} = L_{32}$
$C_{22} = 14.4271\,\text{pF}$
$L_{22} = 1.75573\,\text{nH}$

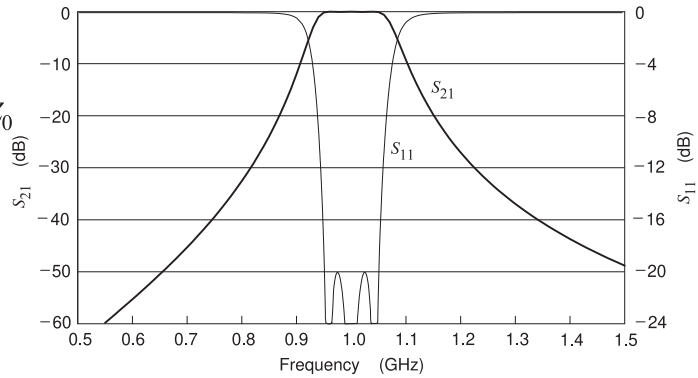
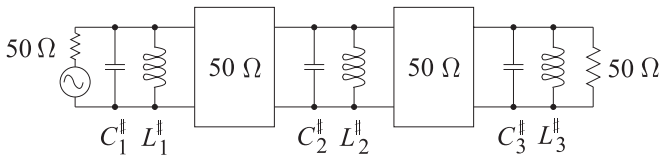


Figure 3-21: A third-order lumped-element Chebyshev bandpass filter with 10% bandwidth, ripple factor of $\varepsilon = 0.1$, and center frequency of 1 GHz. The S parameters are referenced to 1 Ω.



$C_1^\sharp = 27.1066\,\text{pF} = C_3^\sharp \quad L_1^\sharp = 0.934468\,\text{nH} = L_3^\sharp$
$C_2^\sharp = 35.1147\,\text{pF} \quad L_2^\sharp = 0.721359\,\text{nH}$

Figure 3-22: Step 4. Bandpass filter approximation where the inverters are now impedance inverters.

Step 4: Impedance Scaling

The system impedance is now scaled from 1 to 50 Ω, leading to the prototype in Figure 3-22 (derived from Figure 3-20).

Step 5: Conversion of Lumped-Element Resonators

Each lumped-element resonator in Figure 3-22 comprises a capacitor and an inductor. The inductors are particularly difficult to realize efficiently and have high loss. The idea is to replace each lumped resonator (e.g., $C_1^\sharp$ and $L_1^\sharp$ in Figure 3-22) with a network comprising a lumped capacitor and a transmission line stub.

Equivalence of a Capacitively-Loaded Stub and an LC Resonator

Here it will be shown that a capacitively-loaded stub is a broadband approximation of a lumped LC resonator. The resonator equivalence is shown in Figure 3-23. The equivalence is first developed at frequency $f_0 = \omega_0/(2\pi)$ by making sure that the circuits have the same admittance at the

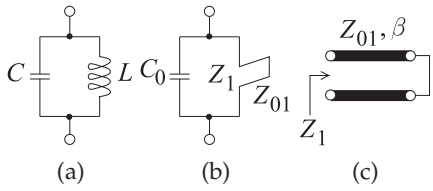


Figure 3-23: Resonator equivalence: (a) lumped resonator resonant at radian frequency ω_0; (b) mixed lumped-distributed resonator containing a short-circuited stub of characteristic impedance Z_0 and resonant at radian frequency ω_r; and (c) short-circuited transmission line stub resonant at radian frequency ω_r. In (b) and (c) Z_{01} is the characteristic impedance of the transmission line and Z_1 is the input impedance of the shorted transmission line.

filter center frequency. The broadband approximation is obtained by also equating the derivatives of the admittances at ω_0.

The input admittance of the parallel LC lumped-element resonator is

$$Y_{\text{in}} = \frac{\jmath\left(\omega^2 LC - 1\right)}{\omega L}, \tag{3.28}$$

and has a radian frequency derivative of

$$\frac{dY_{\text{in}}}{d\omega} = \frac{\jmath\left(\omega^2 LC + 1\right)}{\omega^2 L}. \tag{3.29}$$

Now consider a short-circuited stub that is resonant at radian frequency ω_r; that is, the short-circuited stub is one-quarter wavelength long at ω_r. From Equation (2.100) of [29], the input impedance at ω is

$$Z_1 = \jmath Z_{01} \tan\left(\frac{\pi}{2}\frac{\omega}{\omega_r}\right). \tag{3.30}$$

Thus for the lumped-distributed network in Figure 3-23(b), the input admittance is

$$Y'_{\text{in}} = \frac{\jmath\left[\omega C_o Z_{01} \tan\left(\frac{\pi}{2}\frac{\omega}{\omega_r}\right) - 1\right]}{Z_{01}\tan\left(\frac{\pi}{2}\frac{\omega}{\omega_r}\right)}, \tag{3.31}$$

and its radian frequency derivative is

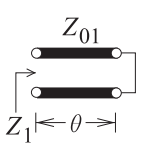
$$\frac{dY'_{\text{in}}}{d\omega} = \jmath\frac{1}{2}\left(\frac{\left\{2C_o Z_{01}\left[\tan\left(\frac{\pi}{2}\frac{\omega}{\omega_r}\right)\right]^2 \omega\omega_r + \pi + \pi\left[\tan\left(\frac{\pi}{2}\frac{\omega}{\omega_r}\right)\right]^2\right\}}{Z_{01}\left[\tan\left(\frac{\pi}{2}\frac{\omega}{\omega_r}\right)\right]^2 \omega_r}\right). \tag{3.32}$$

For the lumped-element and lumped-distributed resonators to be identical at ω_0 and approximating each other at nearby frequencies, Equations (3.28) and (3.31) are equated,

$$Y_{\text{in}}(\omega_0) = Y'_{\text{in}}(\omega_0), \tag{3.33}$$

and Equations (3.29) and (3.32) are also equated,

$$\left.\frac{dY_{\text{in}}}{d\omega}\right|_{\omega=\omega_0} = \left.\frac{dY'_{\text{in}}}{d\omega}\right|_{\omega=\omega_0}. \tag{3.34}$$



At ω_r,
$\theta = 90°, Z_1 = \infty$
At $\omega = \omega_r/2$,
$\theta = 45°, Z_1 = \jmath Z_{01}$

Figure 3-24: Short-circuited transmission line used as a stub resonator in a bandpass filter. The line is resonant at the radian frequency ω_r and so the electrical length of the line is $\theta = 90°$. Using the design choice here, the filter center passband frequency $\omega = \omega_r/2$, and so $\theta = 45°$ and so the input impedance of the shorted stub is $Z_1 = \jmath Z_{01}$.

Specific Design Choice, $\omega_r = 2\omega_0$

At this stage a design choice must be made regarding the relationship of ω_0 and ω_r; that is, the resonant frequency of the stub in relation to the resonant frequency of the resonator (which is also the center frequency of the filter passband). Here a shorted stub is considered that is resonant when it is $\lambda/4$ long. The specific design choice made here is that the resonant frequency of the stub, f_r, is twice the center frequency of the filter passband, i.e. $\omega_r = 2\omega_0$. This is a common design choice. The consequence of this choice is illustrated in Figure 3-24. The frequency f_r is called the **commensurate frequency** to avoid confusion resulting from the resonant frequency of the filter resonators being different from the resonant frequency of stub[2]. The stub design proceeds as follows.

The equivalence expressed by Equations (3.33) and (3.34) is established at the center of the passband (i.e., at ω_0). Now $\omega = \omega_0 = \frac{1}{2}\omega_r$, so Equations (3.28), (3.31), and (3.33) become

$$\frac{\jmath\left[\left(\frac{\omega_r}{2}\right)^2 LC - 1\right]}{\frac{\omega_r}{2}L} = \frac{\jmath\left[\frac{\omega_r}{2}C_0 Z_{01}\tan\left(\frac{\pi}{2}\frac{\omega_r/2}{\omega_r}\right) - 1\right]}{Z_{01}\tan\left(\frac{\pi}{2}\frac{\omega_r/2}{\omega_r}\right)}. \tag{3.35}$$

That is,
$$\frac{\left(\omega_r^2 LC - 4\right)}{2\omega_r L} = \frac{\left[\frac{\omega_r}{2}C_0 Z_0\tan(\pi/4) - 1\right]}{Z_{01}\tan(\pi/4)}. \tag{3.36}$$

Since $\tan\frac{\pi}{4} = 1$,
$$\frac{\left(\omega_r^2 LC - 4\right)}{2\omega_r L} = \frac{\left(\frac{1}{2}\omega_r C_0 Z_{01} - 1\right)}{Z_{01}}, \tag{3.37}$$

and rearranging,
$$C_0 = C - \frac{4}{\omega_r^2 L} + \frac{2}{\omega_r Z_{01}}. \tag{3.38}$$

Another relationship comes from equating derivatives. From Equations (3.29), (3.32), and (3.34), and with $\omega = \frac{1}{2}\omega_r$ and $\tan(\frac{\pi}{2}\frac{\omega}{\omega_r}) = \tan\frac{\pi}{4} = 1$,

$$\frac{\left(\omega_r^2 LC + 4\right)}{\omega_r^2 L} = \frac{1}{2}\frac{\left(2C_0 Z_{01}\frac{1}{2}\omega_r^2 + \pi + \pi\right)}{Z_{01}\frac{1}{2}\omega_r} = \frac{\left(C_0 Z_{01}\omega_r^2 + 2\pi\right)}{Z_{01}\omega_r}. \tag{3.39}$$

Substituting for C_0 from Equation (3.38) and rearranging, the characteristic impedance of the stub is

$$Z_{01} = \frac{\omega_r L}{4}\left(1 + \frac{\pi}{2}\right). \tag{3.40}$$

[2] Of course there are resonant frequencies at every multiple of one-quarter wavelength. The first resonance occurs at f_r, as then the input impedance of the short-circuited stub is an open circuit. The characteristic that establishes resonance is that the input is either an open- or short-circuit and energy is stored.

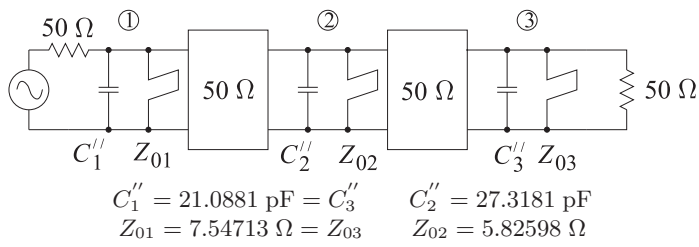


Figure 3-25: Step 5. Bandpass combline filter with impedance inverters. The transmission line stubs present impedances $Z_1 = \jmath Z_{01}$, $Z_2 = \jmath Z_{02}$, and $Z_3 = \jmath Z_{03}$ since the resonant frequencies of the stubs are twice that of the design center frequency, i.e. $f_r = 2f_0$.

In the passband the input impedance of the stub is

$$Z_1 = \jmath Z_{01} \tan(\beta\ell) = \jmath Z_{01} \tan(\pi/4) = \jmath Z_{01}. \tag{3.41}$$

Thus the broadband (≈30%) approximation of the parallel LC resonant circuit is a capacitively-loaded stub, see Figure 3-23(b), of length $\lambda/8$ at the resonant frequency of the LC resonator. The resonant frequency of each stub, i.e. the frequency f_r at which they are $\lambda/4$ long, is twice the design center frequency f_0, i.e. $f_r = 2f_0$. So the concept used here is that the resonant frequency of a stub does not need to be the same as the design center frequency. It is common for all the stubs in a design to have the same resonant frequency f_r and this is specified in the design documentation. In general f_r can have any relationship to f_0 but the most common situation is $f_r = 2f_0$ as used here.

Returning to Step 5

For this filter, the center of the passband is 1 GHz, and so $\omega_0 = 2\pi 10^9$ and $\omega_r = 2\omega_0 = 2\pi(2 \cdot 10^9)$ defines the stub resonant frequency as 2 GHz. The first resonator in Figure 3-22 with $L = 0.934468$ nH and $C = 27.1066$ pF can be replaced by the capacitively-loaded stub in Figure 3-23(b). where, from Equation (3.40),

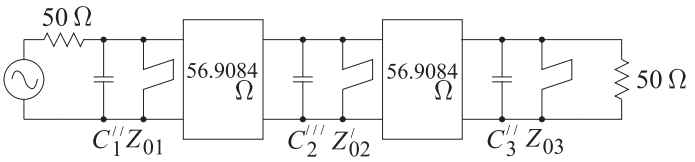
$$Z_{01} = 7.54713\ \Omega, \quad \text{and from Equation (3.38)}, \quad C_0 = 21.0881\ \text{pF}. \tag{3.42}$$

This process is repeated for each lumped-element resonator in Figure 3-22, leading to the prototype shown in Figure 3-25.

Steps 6 and 7: Equating Characteristic Impedances of Stubs

The stubs in Figure 3-25 can be physically realized using transmission line segments. It would be ideal if the impedances of the stubs (i.e., the impedances looking into the stubs) are identical. To achieve this, the method of nodal admittance matrix scaling described in Section 3.3 is used with the impedance inverters multiplied by $\sqrt{Z_{01}/Z_{02}}$ and the admittance, $C_2''//Z_2$, divided by Z_{01}/Z_{02} (C_2'' is divided by Z_{01}/Z_{02}, and Z_2 is multiplied by Z_{01}/Z_{02}). This step scales each inverter impedance to 56.9084 Ω and also sets the characteristic impedance of all the shunt stubs to 7.54713 Ω. The combline filter is now as shown in Figure 3-26.

In Section 2.8.6 it was shown that a good narrowband model of an inverter is a Pi network of stubs. So the inverter in the current filter design, shown in Figure 3-27(a), is modeled by the lumped-element circuit in Figure 3-27(b) and the stub circuit shown in Figure 3-27(c).



$C_1''' = C_3''' = C_2'''' = 21.0881$ pF and $Z_{02}' = 7.54713\ \Omega = Z_{01} = Z_{03}$

Figure 3-26: Step 6. Bandpass combline filter with impedance inverters with $f_r = 2f_0$.

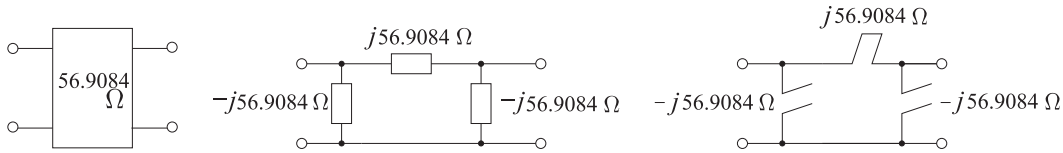


Figure 3-27: Inverter translation used with the combline filter design: (a) impedance inverter; (b) realization as a lumped-element circuit; and (c) realization using stubs resonant at twice the passband center frequency.

Stub Impedance When $f_r = 2f_0$

The most common design choice is to select the resonant (i.e., the commensurate) frequency of the stubs to be twice the center frequency of the design. With $f_r = 2f_0$, Equation (2.114) becomes

$$Z_0 = \frac{1}{J\tan\left(\frac{\pi}{2}\frac{\omega_0}{2\omega_0}\right)} = \frac{1}{J\tan\left(\frac{\pi}{4}\right)} = \frac{1}{J} = K. \tag{3.43}$$

Equation (3.43) defines the characteristic impedance of the transmission line stubs realizing an inverter. Then the elements of the subnetwork in Figure 3-27(b) all have the admittance magnitude

$$J = 1/K = 1/(56.9084\,\Omega) = 17.5721\ \text{mS} \tag{3.44}$$

and the stubs in Figure 3-27(c) have the characteristic impedance

$$Z_0 = \frac{1}{J\tan(\theta)} = \frac{K}{\tan(\theta)} = \left.\frac{K}{\tan\left(\frac{\pi}{2}\frac{\omega}{\omega_r}\right)}\right|_{\omega=\omega_0} = 56.9084\,\Omega. \tag{3.45}$$

At this stage there are several pairs of stubs in parallel. Figure 3-28 illustrates how a pair of parallel stubs can be replaced by a single stub. Now the prototype is as shown in Figure 3-29.

Step 8: Scaling Characteristic Impedances of Stubs

The filter will be realized using coupled microstrip lines and a general statement can be made about how the stubs in the prototype in Figure 3-29 correspond to the physical structure of the coupled lines. The stubs with characteristic impedances Z_{01}', Z_{02}'', and Z_{03}' are largely realized by the properties of the individual microstrip lines. A low Z_{01}', for example, would

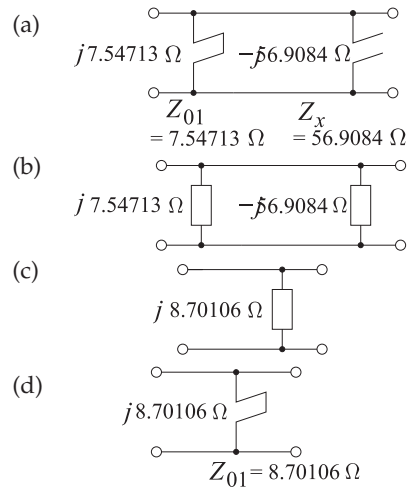


Figure 3-28: Step 6b. Procedure combining two parallel stubs to obtain one stub: (a) two parallel stubs; (b) represented as parallel impedances; (c) as one impedance; and (d) final representation as one stub. Note that the stubs must have the same commensurate frequency, i.e. they must have the same electrical length.

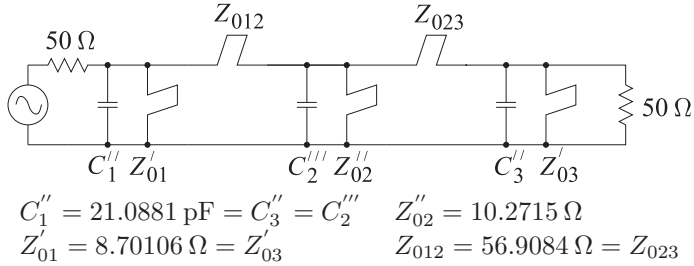


Figure 3-29: Step 7: Bandpass filter approximation where Z'_{01}, Z''_{02}, Z'_{03}, Z_{012}, and Z_{023} are the characteristic impedances of the stubs with commensurate frequency $f_r = 2f_0$.

require a wide microstrip line. The stubs with characteristic impedances Z_{012} and Z_{023} each correspond to coupling between pairs of microstrip lines with high impedances corresponding to low-level coupling and widely spaced microstrip lines.

In Figure 3-29 the short-circuited shunt stubs have characteristic impedances of $8.7\ \Omega$ and $10.3\ \Omega$, which are too low, and will result in wide microstrip lines.[3] These need to be scaled to a higher impedance level to raise the characteristic impedances of the short-circuited stubs to an acceptable value. Note that the characteristic impedance of the middle shunt short-circuited stub can be raised to $80\ \Omega$ if the system impedance is raised from $50\ \Omega$ to $389.426\ \Omega$. Doing this leads to the element values shown in the bandpass circuit of Figure 3-30.

Step 9: Scaling to a 50 Ω System Impedance

The previous step raised the system impedance (the required source and load impedances) to an unreasonably high level ($389.426\ \Omega$). For the filter to operate in a $50\ \Omega$ system, input and output impedance inverters are required to scale the source and load back to $50\ \Omega$. The resulting circuit is shown in

[3] What is reasonable for the width of a microstrip line is subjective and based on experience. With a substrate having $\varepsilon_r = 10$, the width required for a characteristic impedance, Z_0, of 30–80 Ω is reasonable. A line with $Z_0 < 30\ \Omega$ will be very wide, use too much area, and potentially lead to multi-moding. A line with $Z_0 > 80\ \Omega$ will be narrow, close to the maximum Z_0 that can be realized in a technology, have manufacturing tolerance issues, and have large radiation. The range can be extended to 20–100 Ω but typically with degraded performance. Also see Example 3.4 of [29] for a microstrip design guideline.

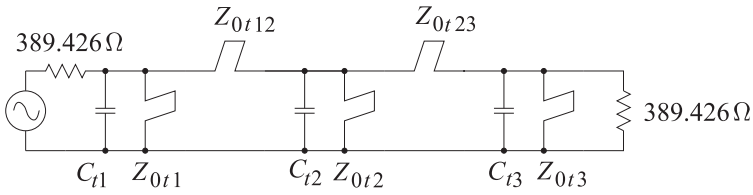


Figure 3-30: Step 8: Bandpass filter approximation.

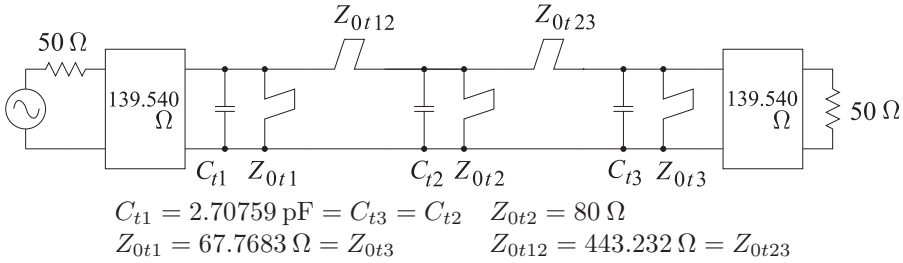


Figure 3-31: Step 9: Bandpass filter approximation.

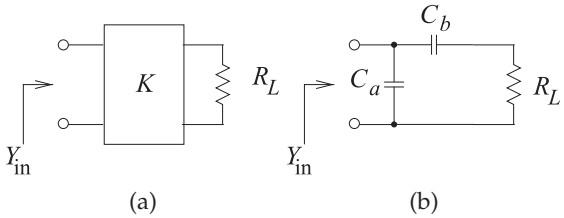


Figure 3-32: Inverter equivalence: (a) resistively-terminated impedance inverter; and (b) its equivalent capacitor network. Here $Z_{\text{in}} = 1/Y_{\text{in}} = 389.426\ \Omega$, $R_L = 50\ \Omega$, and $K = 139.540\ \Omega$.

Figure 3-31 with input and output inverters of impedance $\sqrt{50 \times 389.426} = 139.540\,\Omega$.

3.4.1 Realization of the Input/Output Inverters

At this stage the design is as shown in Figure 3-31, where the inverters at the input and output remain to be synthesized. There are many possible realizations of the inverters. Focusing on just the output inverter with a load R_L, as shown in Figure 3-32(a), it will be shown here how this can be realized using the capacitive network of Figure 3-32(b).

The input admittance of the inverter plus load resistor of Figure 3-32(a) is

$$Y_{\text{in}} = \frac{R_L}{K^2}, \tag{3.46}$$

and the input admittance of the circuit in Figure 3-32(b) is

$$Y_{\text{in}} = sC_a + \frac{1}{1/(sC_b) + R_L}. \tag{3.47}$$

Thus, equating the real parts of Equations (3.46) and (3.47),

$$\Re\,(Y_{\text{in}}) = \frac{R_L\omega^2 C_b^2}{R_L^2\omega^2 C_b^2 + 1} = \frac{R_L}{K^2}, \tag{3.48}$$

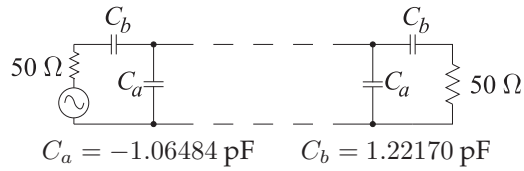


Figure 3-33: The external inverters of the prototype shown in Figure 3-31 replaced by capacitive networks.

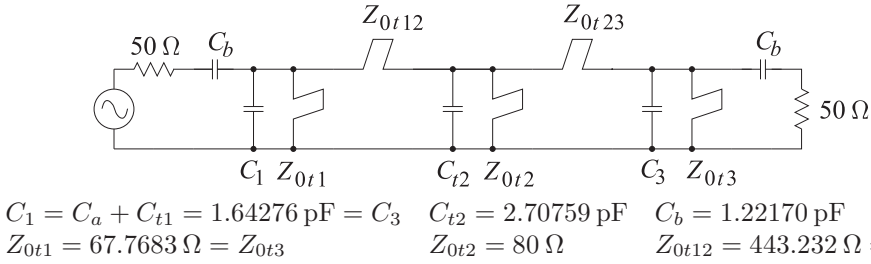


Figure 3-34: Step 10: The bandpass filter approximation combining the capacitive equivalent of the inverters with the first and last capacitors in the circuit of Figure 3-31.

and equating the imaginary parts yields

$$\Im\,(Y_{\text{in}}) = \omega\left[C_a + \frac{C_b}{1+(\omega C_b R_L)^2}\right] = 0. \tag{3.49}$$

Thus $$C_b = \frac{1}{\omega K\sqrt{1 - R_L^2/K^2}}, \quad C_a = \frac{-C_b}{1+(\omega C_b R_L)^2}, \tag{3.50}$$

and the inverters shown in Figure 3-31 can be replaced by the capacitive networks shown in Figure 3-33. This capacitive network is not a general replacement of an inverter. It only works when the inverter is terminated in a resistor. Note that C_a is negative, and this is absorbed in the first and last resonators so that the final lumped-distributed realization is as shown in Figure 3-34. This is the electrical design of the bandpass filter in a form that can be realized using the combline connection of coupled lines.

3.4.2 Implementation

The filter designed in the previous section can be implemented using three parallel microstrip lines, as shown in Figure 3-35(a), where the three capacitors of the electrical design in Figure 3-34 appear directly in Figure 3-35(a). The double Pi connection of stubs shown in Figure 3-34, and again in Figure 3-36(a), becomes the three coupled lines shown in Figure 3-36(b).

Derivation of Physical Dimensions

The task now is to determine the dimensions of the three parallel microstrip lines. The dimensions to be determined are shown in Figure 3-36(b) and are the length of the lines, L, the widths of the three lines, w_1, w_2, and w_3, and the gaps s_1 and s_2. The approach is to consider that the three parallel lines form two coupled-line pairs with each pair of coupled lines implementing a Pi network of stubs. The physical design procedure is outlined in Figure 3-36 with one of the lines shared. This, of course, is an approximation but a fairly

accurate one. The physical synthesis of a pair of coupled lines is outlined in Figure 3-37 where the parameters of the Pi model, Figure 3-37(a), lead to the parameters of the transmission lines in the combline model, Figure 3-37(b). This then leads to the parameters defining the pair of coupled lines, Figure 3-37(c), in particular the even and odd-mode characteristic impedances, Z_{0e} and Z_{0o} and the system impedance Z_{0S}.

Using the quantities in Figure 3-37(a) and rearranging Equations (3.14)–(3.16), the parameters of the coupled-line model in Figure 3-37(b) are obtained as follows:

$$n = 1 + \frac{Z_{012}}{Z_{011}}, \quad Z_{02} = nZ_{012} \quad \text{and} \quad Z_{01} = n\left(\frac{Z_{011}Z_{022}}{Z_{011} + Z_{022} + Z_{012}}\right). \quad (3.51)$$

For the left pair (and also for the right pair because of symmetry)) of coupled lines in Figure 3-36(d), $Z_{012} = Z_{0t12} = 443.232\ \Omega$, $Z_{011} = Z_{0t1} = 67.7683\ \Omega$,

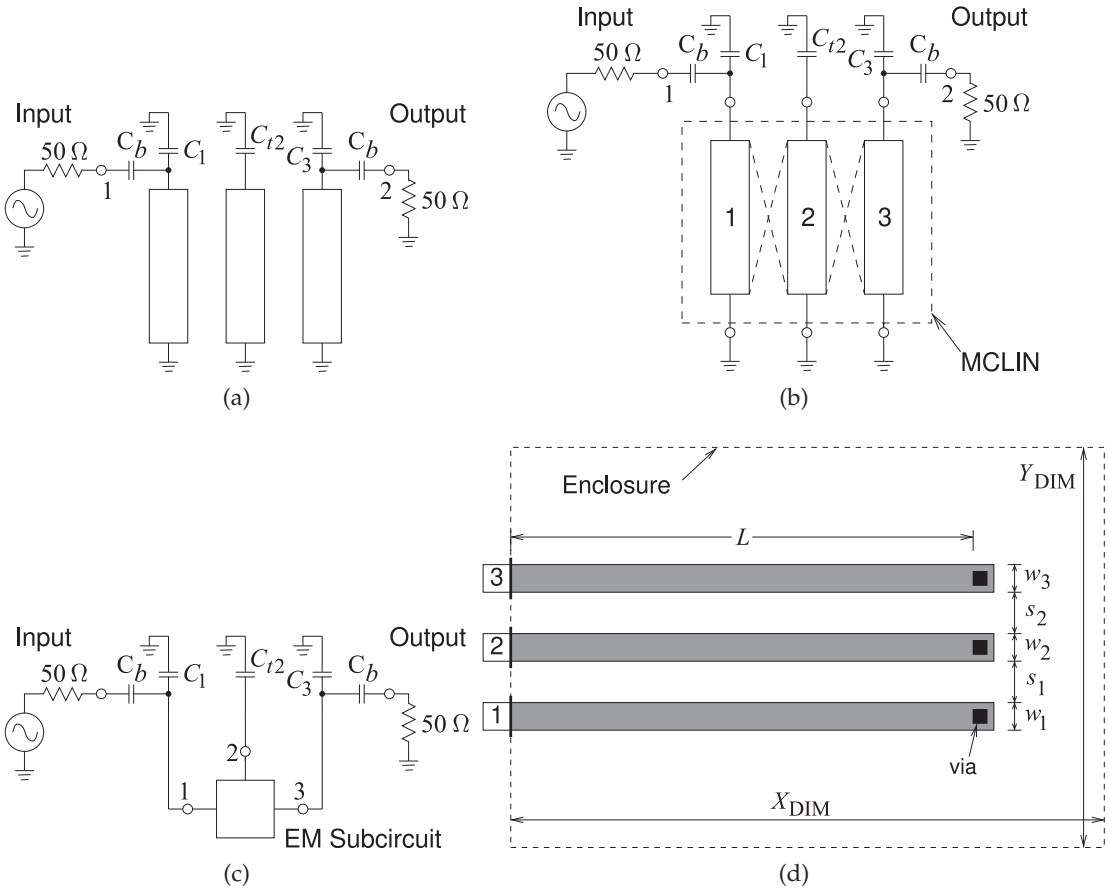


Figure 3-35: Physical layout of the combline bandpass filter designed in Section 3.4.1: (a) with discrete capacitors; (b) with a microstrip coupled transmission line element (MCLIN); (c) with an EM subcircuit block; and (d) the EM subcircuit block with layout to be simulated in an EM simulator. Details: 6 μm gold metalization, 635 μm thick alumina substrate with $\varepsilon_r = 10$, 400 μm × 400 μm tantalum vias, and the EM enclosure has perfectly conducting walls with $X_{\text{DIM}} = 22$ mm, $Y_{\text{DIM}} = 20$ mm, and height = 5.635 mm.

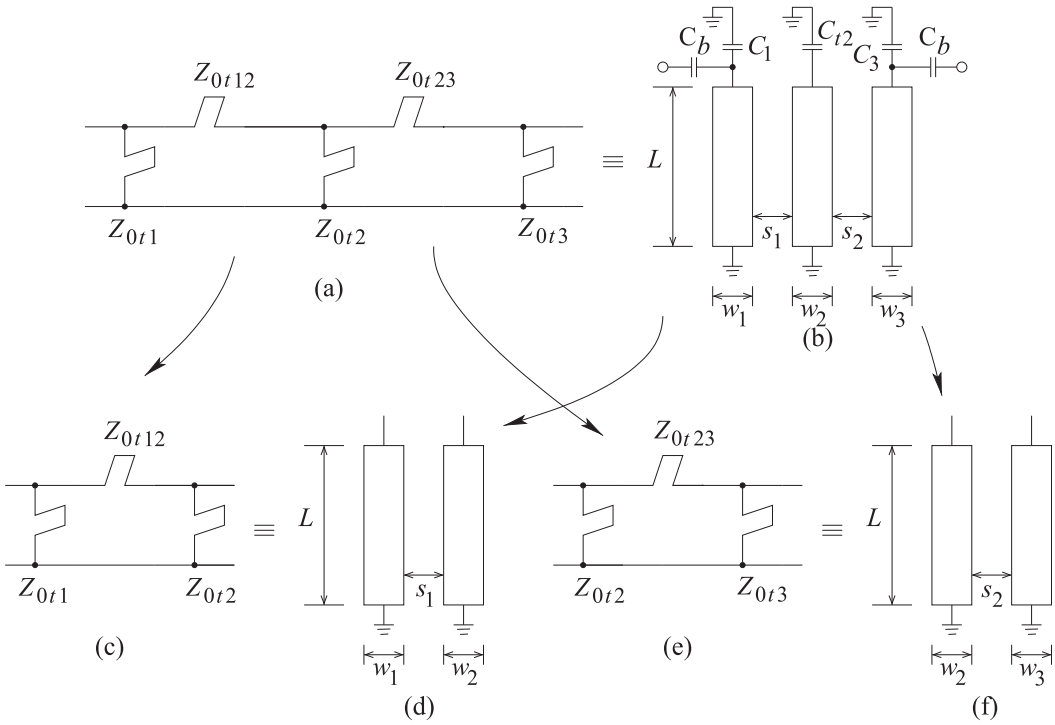


Figure 3-36: Physical design of the third-order combline bandpass filter. The double Pi network in (a) is partitioned into two Pi networks in (c) and (e). Similarly the three coupled microstrip lines in (b) are partitioned into two pairs of coupled lines in (d) and (f). Then physical design of the filter involves transforming the electrical design in (c) into the physical design in (d). Similarly the circuit in (e) is transformed into the layout in (f). With a shared central stub, the electrical circuit in (a) becomes the three coupled-line realization in (b).

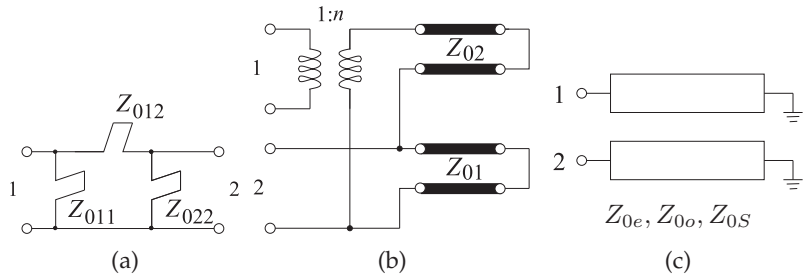


Figure 3-37: Equivalent circuits for a combline section (from Figure 3-13).

$Z_{022} = Z_{0t2} = 80\ \Omega$ and so[4]

$$n = 1 + \frac{Z_{012}}{Z_{011}} = 7.540, \quad Z_{02} = 3342\,\Omega, \quad \text{and} \quad Z_{01} = 69.17\,\Omega. \tag{3.52}$$

[4] Note that the precision has been reduced. It was necessary to retain high precision during the electrical synthesis but the physical dimensions do not need the same precision.

The physical parameters of the coupled line pair in Figure 3-37(c) are derived using the results in Section 5.9.6 of [29]. Thus

$$K = 1/n = 0.1326. \tag{3.53}$$

From Equations (5.197) and (5.198) of [29] two estimates of the system impedance, Z_{0S}, are obtained:

$$Z_{0S,(3.54)} = Z_{01}\sqrt{1-K^2} = 68.56\,\Omega \tag{3.54}$$

$$Z_{0S,(3.55)} = Z_{02}\frac{K^2}{\sqrt{1-K^2}} = 59.30\,\Omega. \tag{3.55}$$

These two values of Z_{0S} are different and this is because the Pi arrangement of stubs is not symmetrical (i.e., in Figure 3-37(a) $(Z_{011} = 67.7\,\Omega)$ $/\!\!=$ $Z_{022} = 80\,\Omega$)).[5] The assumption behind the development of Equations (3.3) and (3.4) are that the pair of coupled lines is symmetrical. The only reasonable choice is to use the geometric mean of the two values. So

$$Z_{0S} = Z_{0S,\text{mean}} = \sqrt{Z_{0S,(3.54)}Z_{0S,(3.55)}} = 63.77\,\Omega. \tag{3.56}$$

Then, from Equation (5.202) of [29],

$$Z_{0o} = Z_{0S}\sqrt{\frac{n-1}{n+1}} = 55.80\,\Omega, \quad Z_{0e} = \frac{Z_{0S}^2}{Z_{0o}} = 72.87\ \ \Omega, \text{ and } \ \frac{Z_{0e}}{Z_{0o}} = 1.306. \tag{3.57}$$

The next step is to choose a substrate and realize the physical dimensions of the coupled lines. It is important that there be dimensional stability and that there be no multimoding. The best substrate for dimensional stability is a hard substrate like alumina, sapphire, or glass. Soft substrates like FR4 or teflon are too variable to be useful for filters. Alumina is particularly attractive as it is very hard and dimensions are reproducible. Alumina shrinks when it is processed but the amount of shrinkage is well controlled and reproducible. So once a design has been physically tuned, typically by drilling out part of the dielectric to change the effective permittivity, filters can be reliably reproduced and require only a small amount of adjustment. Alumina is polished to a well defined thickness and with a relative permittivity of around 10 a 50-Ω microstrip line has a width approximately equal to the thickness of the substrate. Having equal dimensions is a good thing to have for manufacturability. The thickness of the substrate must be decided on at this stage in design. A thicker substrate is less likely to crack during handling however a thicker substrate is more likely to support multimoding. Here a $(h =)$ 635 μm-thick alumina substrate with $\varepsilon_r = 10$ is chosen and calculations (not shown here) indicate that there will not be multimoding at the filter's operating frequencies.

Physical dimensions could be derived by iteratively solving for Z_{0e} and Z_{0o} using the formulas in Section 5.6 of [29]. As an approximation, here the

[5] Thus error is being introduced at this step. In reality the final manufactured filter will need to be tuned anyway as the filter will require tolerances of about 0.1% or better and the best manufactured dimensional tolerance is usually about 1%. Also material parameters are usually not known to better than 1%. So this is an error that must be accepted.

physical dimensions are derived from Table 5-3 of [29]. Table 5-3 of [29] is for a system impedance, Z_{0S}, of $50\ \Omega$ and not the $63.77\ \Omega$ found here. Thus there will be an error, but it can be expected that there will be error in any case and these can be resolved using optimization in a simulator. From Table 5-3 of [29], the initial physical dimensions of the pairs of lines are

$$u = 0.93, w = uh = 591\ \mu\text{m}, \quad \text{thus with rounding } \ w_1 = w_2 = w_3 = 600\ \mu\text{m} \tag{3.58}$$

$$g = 1.0, s = gh = 635\ \mu\text{m}, \quad \text{thus with rounding } \ s_1 = s_2 = 650\ \mu\text{m} \tag{3.59}$$

The rounding has been done as for EM simulation the lines will be laid out on a regular grid and a $50\ \mu\text{m}$ grid is reasonable.

Scaling of Line Width

At this stage the error made with the characteristic impedance of the microstrip line should be considered. While any error can be corrected when optimizing the final design, making some correction at this stage will reduce the final adjustments needed. Three system impedances have been derived, $Z_{0S,(3.54)} = 68.56\ \Omega$, $Z_{0S,(3.55)} = 59.30\ \Omega$, and the geometric mean $Z_{0S,\text{mean}} = 63.77\ \Omega$. This divergence is one source of error. Another source of error is that a table for a $50\ \Omega$ system impedance was used to determine the coupled-line parameters. Note that the electrical design is exact and the tuning of the physical design will attempt to match the response of the physical circuit to that of the electrical design. However some initial corrections will reduce the amount of tuning that needs to be done.

Scaling for the higher actual system impedance ($63.77\ \Omega$ versus $50\ \Omega$) means using higher impedance microstrip lines (which will be narrower) and higher coupling impedance (requiring a greater separation of the lines). There are three possible characteristic impedance to scale to and a design decision needs to be made. Should the lines be scaled from $50\ \Omega$ to $54.2\ \Omega$, $68.6\ \Omega$, or $61.9\ \Omega$? The most conservative approach is to scale from $50\ \Omega$ to $54.2\ \Omega$, a factor of 1.08. Here only adjustment of the line widths will be considered and two correction approaches will be described. The first is based on the rule of thumb described in Example 3.4 of [29] which indicated that the characteristic impedance of the microstrip lines is approximately proportional to $1/\sqrt{w}$. So the line width should be reduced by a factor of $1.08^2 = 1.17$ to shift the characteristic impedance. Thus the line width should be reduced from $600\ \mu\text{m}$ to about $600\ \mu\text{m}/1.17 = 512\ \mu\text{m}$ or $500\ \mu\text{m}$ after rounding.

A second approach to correction is based on Table 3-3 of [29]. For $\varepsilon = 10$ and using the more conservative system impedance (i.e. $55.8\ \Omega$ which is closest to $50\ \Omega$), a better choice for u is to replace $u = 0.93$ by $u = 0.78$ and, after rounding, $w_1 = w_2 = w_3 = 500\ \mu\text{m}$. This cannot be expected to completely correct for the error, but it should be closer than the choice of $600\ \mu\text{m}$.

Scaling the system impedance should also require that s be increased but there is not a simple guideline to follow and so this will be skipped. Another reason for needing to increase s is that there is more coupling in our coupled line system than that predicted by considering two pairs of coupled lines. For example, there is coupling from lines 1 directly to line 3. To reduce the coupling s will need to be increased, perhaps significantly. Another source

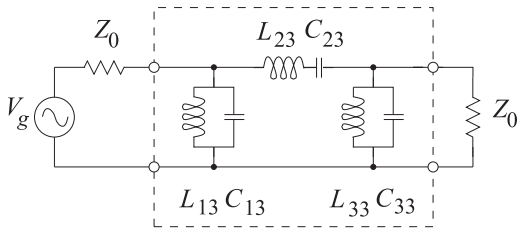


Figure 3-38: Lumped-element bandpass filter in a 50 Ω system. (Derived by multiplying the impedances in Figure 3-21 by 50.)
$Z_0 = 50\ \Omega$,
$C_{13} = C_{33} = 27.107$ pF,
$L_{13} = L_{33} = 934.47$ pH,
$C_{23} = 288.54$ fF, $L_{23} = 87.787$ nH.

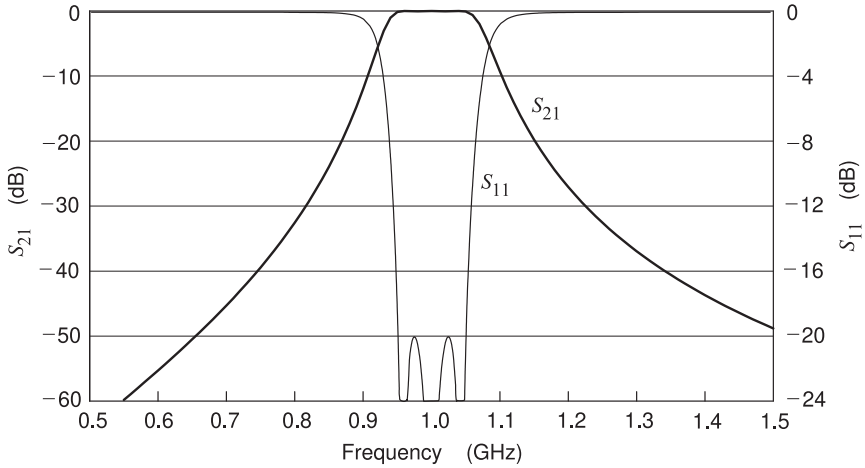


Figure 3-39: Return loss and insertion loss for the filter modeled using the lumped-element model.

of increased coupling is that the phase velocities of the even and odd modes differ. This tends to increase coupling. The adjustment of s is left to tuning of the physical design in an EM simulator.

Line Lengths

The stubs are one-eighth wavelength long at the center frequency of the filter. Thus the coupled lines are also one-eighth wavelength long. From Table 5-3 of [29], the effective relative permittivity of the even mode is $\varepsilon_{ee} = 7.24$ and for the odd mode is $\varepsilon_{eo} = 5.95$. These are quite different so the best that can be done is to use the geometric mean. Thus the effective relative permittivity is $\varepsilon_e = \sqrt{\varepsilon_{ee}\varepsilon_{eo}} = 6.56$. Thus at 1 GHz the line length $L = \lambda_8/8 = \lambda_0/\sqrt{\varepsilon_e} = (30\ \text{cm})/(8\sqrt{6.56}) = 14.64$ mm (14.65 mm with rounding). The use of a single effective permittivity in determining the lengths of the lines is perhaps the largest error and will result in the resonant frequency of each of the resonators, consisting of a capacitor (either C_1, C_{t2}, or C_3) and the $\lambda/8$-long line, being shifted from 1 GHz. The resonators can be re-centered by adjusting the capacitors. Adjusting the capacitance values also corrects for error in the widths of the lines since the main effect of an error in the width of a line is an error in its characteristic impedance and hence the input impedance of the resonator stub.

Microwave Circuit Simulation

Before simulating the combline filter the lumped-element bandpass filter will be considered to provide a benchmark. The 50-Ω lumped-element bandpass filter is shown in Figure 3-38. The responses of this filter is shown in Figures 3-39, 3-40, and 3-41. These responses provide a reference as the coupled-line

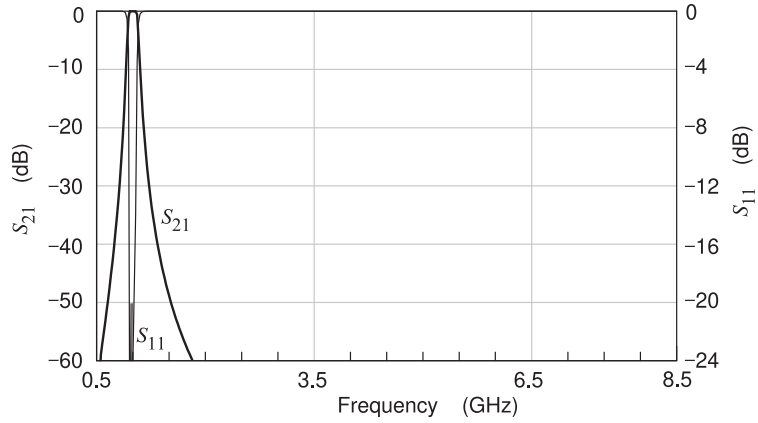


Figure 3-40: Wideband insertion loss and return loss for the filter modeled using the lumped-element model.

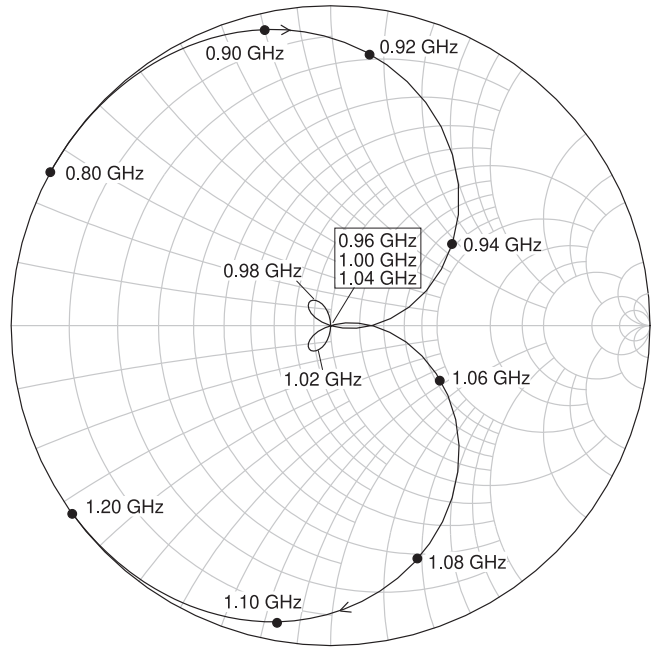


Figure 3-41: Plot of S_{11} on a Smith chart for the lumped-element bandpass filter. The zeros of the S_{11} response, and hence the poles of the S_{21} response, are at 0.96, 1.00, and 1.04 GHz. The looping near the origin is characteristic of Chebyshev filters.

filter design is optimized.

Figure 3-35(a) shows the combline filter to be modeled in a microwave circuit simulator. In a microwave simulator the filter can be modeled using a coupled microstrip element (known as the MCLIN element) or using EM simulation of an actual layout. The first set of results that will be presented uses the coupled microstrip element line model, the MCLIN element, and the filter model is as shown in Figure 3-35(b). The results of the circuit simulation are shown in Figure 3-42. The curves (a) are responses for a gap of $s = 650$ μm. The passband is too wide and this is reduced by increasing s to 1150 μm yielding curves (b). The effect of loss is seen with S_{21} less than 0 dB in the passband. The S_{21} response has a slight slope down in the passband and the S_{11} response does not show the three zeros seen in the lumped-element response, see Figure 3-39.

Plotting the S_{11} response of the MCLIN-based filter on a Smith chart, see Figure 3-43, provides more insight into what is happening. This should be contrasted to the S_{11} response of the lumped-element filter shown in Figure

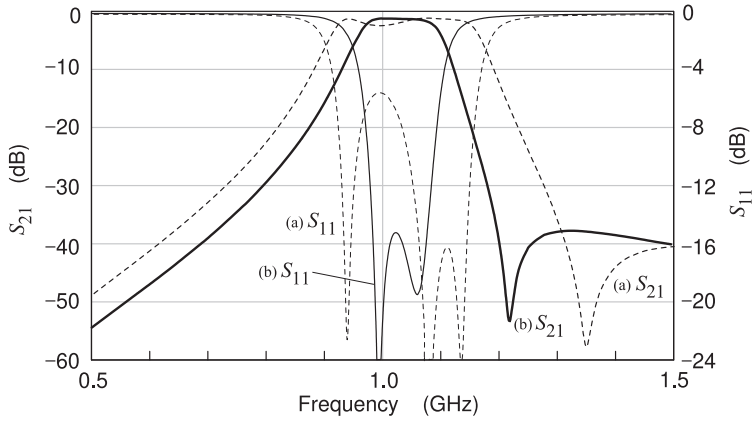


Figure 3-42: Insertion loss and return loss for the filter modeled using the MCLIN element with a line length of $L = 14{,}650$ µm and line widths $w_1 = w_2 = w_3 = 500$ µm. $C_1 = C_3 = 1.6428$ pF, $C_{t2} = 2.7076$ pF, and $C_b = 1.2217$ pF. For curves (a) $s_1 = s_2 = 650$ µm and for curves (b) $s_1 = s_2 = 1150$ µm.

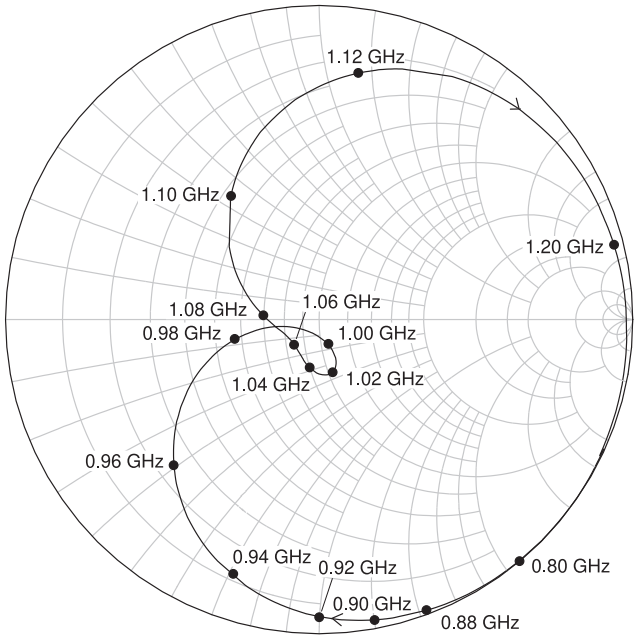
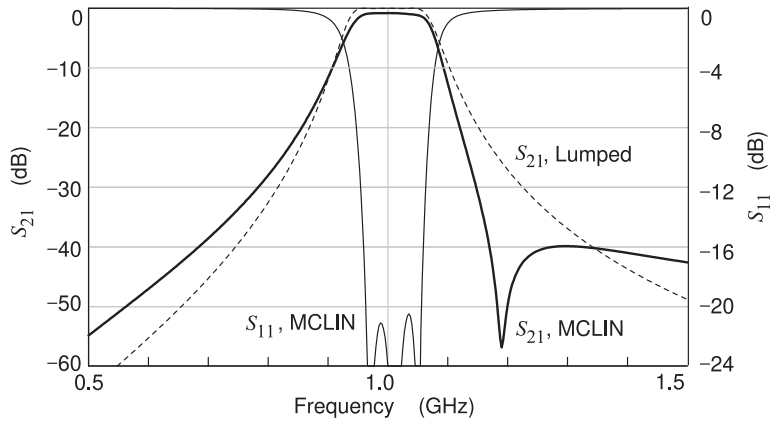


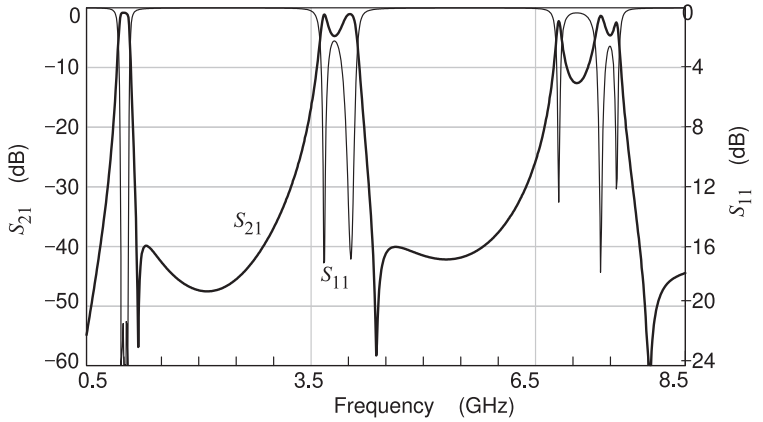
Figure 3-43: Plot of S_{11} on a Smith chart for filter modeled using the MCLIN element and gaps of $s_1 = s_2 = 1150$ µm and a line length of $L = 14{,}550$ µm and line widths $w_1 = w_2 = w_3 = 500$ µm. $C_1 = C_3 = 1.6428$ pF, $C_{t2} = 2.7076$ pF, and $C_b = 1.2217$ pF.

3-41. Recall that the lumped-element filter is the ideal electrical design. The looping of S_{11} at the origin in Figure 3-41 indicates that the ideal electrical response has three zeros as the locus of S_{11} passes through the origin three times. This behavior corresponds to the poles of the transmission response but these poles cannot be seen in the S_{21} response. Consequently in optimizing a filter the designer focuses on the S_{11} response.

Again consider the S_{11} responses of the two filters plotted on Smith charts (i.e. Figures 3-41 and 3-43). The S_{11} response of the lumped-element filter has two small loops corresponding to peaks in the in-band $|S_{11}|$ response seen in the rectangular plot of Figure 3-39. In contrast, the locus of S_{11} of the MCLIN-based filter does not go through the origin, see Figure 3-43, and there is one loop although the hint of a second loop is seen at 1.06 GHz. The performance is close to that required but optimization is necessary. In manually optimizing, also called tuning, the design of the MCLIN-based filter, both the rectangular and Smith chart plots should be overlaid with the lumped-element response. By adjusting C_1, C_{t2}, and C_3 (by 6%, 16%,



(a) Narrowband response



(b) Wideband response

Figure 3-44: Insertion loss and return loss for the filter modeled using the MCLIN element and gaps of $s_1 = s_2 = 1150$ μm and a line length of $L = 14{,}550$ μm and line widths $w_1 = w_2 = w_3 = 500$ μm. $C_1 = C_3 = 1.9076$ pF, $C_{t2} = 2.8748$ pF, and $C_b = 1.2217$ pF.

and 6% respectively), the resonant frequencies of the resonators are centered at 1 GHz and the filter response is also centered at 1 GHz[6]. The result of this tuning is shown in Figures 3-44 and 3-45 indicating that the required performance has been obtained.

In Figure 3-44(a) the MCLIN-based filter response is compared to the lumped-element response. The bandwidths of the MCLIN and lumped-element implementations are closely matched. The MCLIN implementation has the steep skirts resulting from the Chebyshev prototype. The Smith chart plot of S_{11} of the MCLIN-based filter is shown in Figure 3-45. The loops seen with the lumped-element response, in Figure 3-41, are seen but now there is an additional rotation with respect to frequency. This is because the MCLIN-based filter has additional transmission-line delays and hence frequency-dependent phase shift. Also the S_{11} response in Figure 3-45 is rotated by $180°$ from the S_{11} locus for the lumped-element filter (see Figure 3-41). This is because of the additional inverter at the input of the distributed filter design.

Some gross features are seen in the combline response that are not seen in the ideal electrical response. Referring to Figure 3-44(a), these include the additional zero seen in the S_{21} response at 1.2 GHz, the steeper skirt above

[6] Approximately, the resonant frequency of a resonator is inversely proportional to $\sqrt{C}$. So increasing the capacitances will reduce the center frequency of the filter.

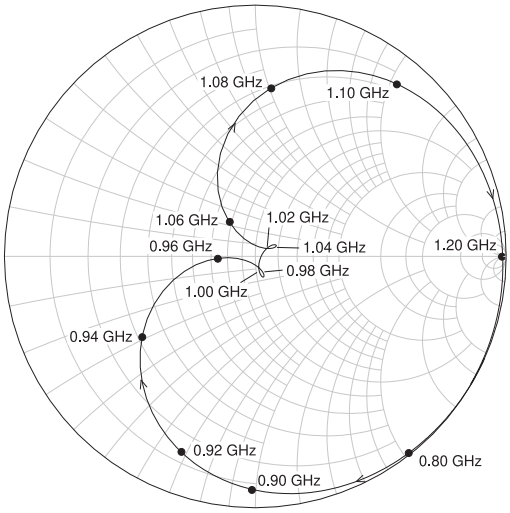
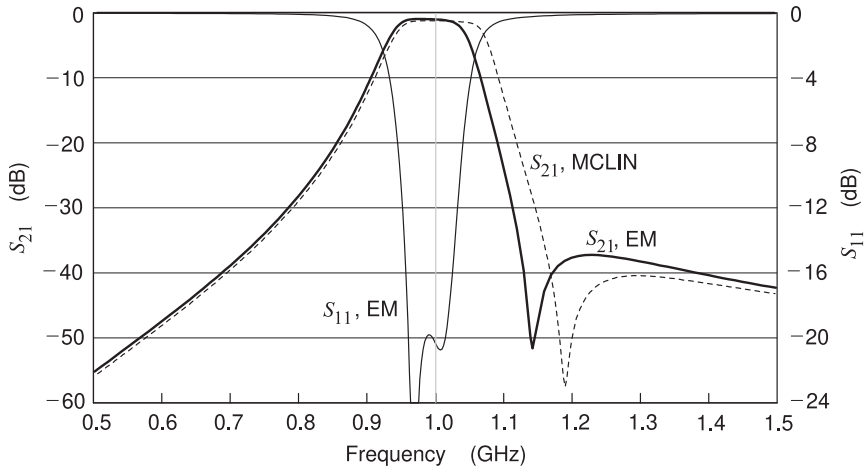


Figure 3-45: Plot of S_{11} on a Smith chart for filter modeled using the MCLIN element and gaps of $s_1 = s_2 = 1150$ µm and a line length of $L = 14{,}550$ µm and line widths $w_1 = w_2 = w_3 = 500$ µm. $C_1 = C_3 = 1.9076$ pF, $C_2 = 2.8748$ pF, and $C_b = 1.2217$ pF.

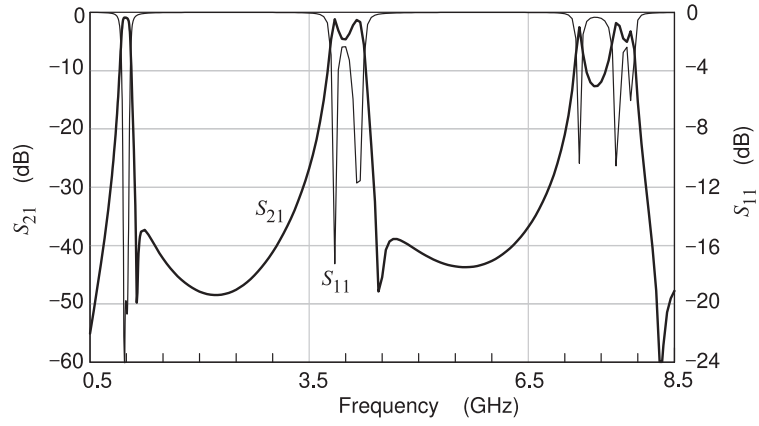
the 1 GHz passband, and the less-steep skirt below the passband. These are related. These features are common to PCL designs that use a combline. The origin of this behavior is the additional transmission path directly from the first line to the last line of the filter, the 1-3 coupling. During design the only transmission path considered was from line 1 to line 2, and then from line 2 to line 3, the 1-2-3 coupling. Apparently, at 1.2 GHz the 1-3 and 1-2-3 signals have the same magnitude but opposite phase and so cancel producing the S_{21} zero. Above the passband the signals on the two paths destructively interfere causing the steeper skirt. Below the passband the 1-3 and 1-2-3 signals constructively interfere and reduce the steepness of the filter skirt. The additional zero can be exploited to significantly reduce the level of a particular signal such as a local oscillator or a signal in a neighboring communication band. If the reduced filter skirt below the passband is not acceptable, then another arrangement of coupled lines should be used.

Figure 3-44(b) is a wideband response of the MCLIN-based filter and shows the spurious passbands that occur with most transmission-line based networks. In this case the properties of the lines are the same at $\lambda/8$, $5\lambda/8$, and $9\lambda/8$. So, by just considering the transmission lines alone it would be expected that there would be spurious passbands at 5 GHz and 9 GHz. The spurious passbands are lower because the impedances of the capacitances reduce at the higher frequencies, resulting in the passbands at four times and seven times the center frequency of the filter. In practice, a simple lowpass filter eliminates the spurious passbands. Often parasitics in the system do this without specific circuitry. Also, appropriate choice of matching networks can provide adequate elimination of the spurious passbands.

Before finalizing the filter layout, a more accurate EM simulation is required instead of using the MCLIN element. The EM simulation captures more subtle effects than can be included in the MCLIN model. The layout of the parallel coupled lines is shown in Figure 3-35(d) and the connection to incorporate these in a complete circuit simulation is shown in Figure 3-35(c). The results of the EM-based simulation are shown in Figures 3-46 and 3-47. It can be seen that the bandwidth of the filter reduces. Additional optimization would result in the EM-based filter response more closely matching the ideal electrical response.



(a) Narrowband response



(b) Wideband response

Figure 3-46: Return and insertion loss for the filter modeled using EM modeling and gaps of $s_1 = s_2 = 1150$ μm and a line length of $L = 14{,}550$ μm and line widths $w_1 = w_2 = w_3 = 500$ μm. $C_1 = C_3 = 1.9076$ pF, $C_{t2} = 2.8748$ pF, and $C_b = 1.2217$ pF.

3.4.3 Alternative Combline Filter Layouts

Up to a few gigahertz surface-mount capacitors can be used in the implementation of the capacitively-loaded stubs. Above that other arrangements should be considered. There are several layout variations for the combline filter. The main variations relate to replacing the lumped-element capacitors by gap capacitors and alternative realizations of the input and output inverters. Several variations are shown in Figure 3-48. The layout in Figure 3-48(a) implements the capacitors using gap capacitors. At 1 GHz it is difficult to realize the required capacitance using gap capacitors, then the interdigitated capacitors shown in Figure 3-48(b) can be used to obtain higher capacitance values. Alternatively a coupled-line section can be used to implement the input and output inverters as shown in Figure 3-48(c).

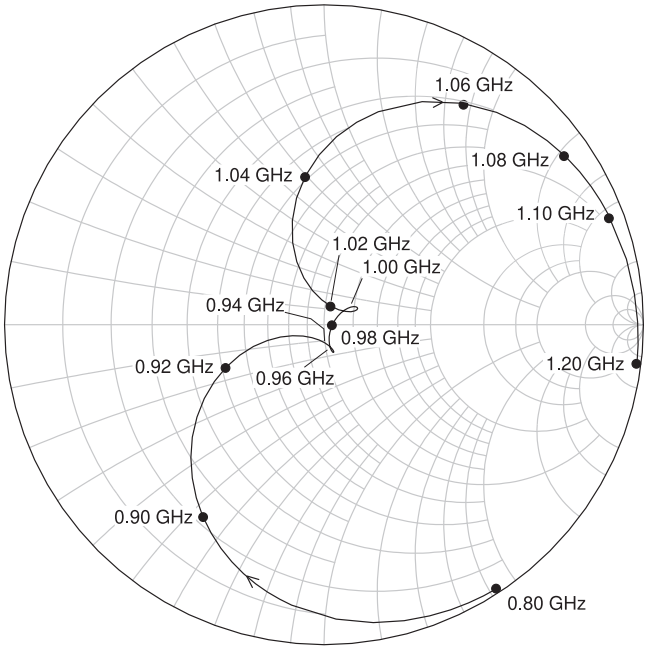


Figure 3-47: Plot of S_{11} on a Smith chart for filter modeled using EM modeling and gaps of $s_1 = s_2 = 1150$ μm, line length $L = 14{,}550$ μm, line widths $w_1 = w_3 = w_2 = 500$ μm. $C_1 = C_3 = 1.9076$ pF, $C_2 = 2.8748$ pF, and $C_b = 1.2217$ pF.

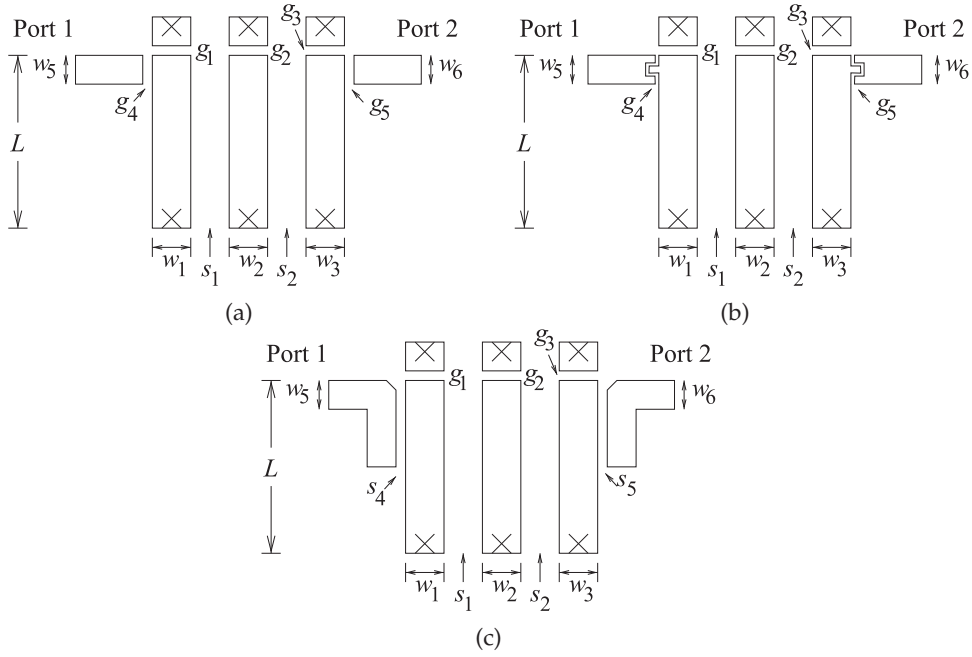


Figure 3-48: Physical layouts of the combline bandpass filter: (a) with gap capacitors having gaps of g_1, g_1, ..., g_5,; (b) with interdigitated capacitors to obtain higher capacitance; and (c) with coupled lines (with separations s_4 and s_5) realizing the input and output inverters.

3.5 Parallel Coupled-Line Filters in an Inhomogeneous Medium

A complexity arises when a parallel coupled-line filter is realized in an inhomogeneous transmission medium due to the inequality of the even- and odd-mode phase velocities. This effect can be investigated by utilizing the full $ABCD$ matrix (see Equation (5.184) of [29].) of each PCL section in a filter for a known even- to odd-mode velocity ratio. The inhomogeneous medium can result in additional spurious passbands centered approximately at even multiples of f_0 [4, 32–45]. This provides all the more reason to perform an accurate full EM simulation following filter synthesis.

3.6 Summary

Filters using parallel coupled-line (PCL) segments are an important class of microwave filters. There are various ways a pair of coupled lines, a four-port, can be configured using shorts and opens to realize a two-port network. Many of these PCL configurations have desirable frequency selectivity characteristics. That is, they inherently have the bandpass, lowpass, highpass, or bandstop characteristics desired of a filter. In particular, the various configurations often have very sharp responses. One drawback, however, is that they have spurious passbands that derive from a line and its counterpart that is one-half wavelength longer having the same input reflection coefficient. These spurious passbands can often be pushed up in frequency by setting, in the case of a bandpass filter, the resonant frequency of the transmission line resonators above the center frequency of the filter being developed.

The synthesis of PCL filters follows the general microwave design philosophy of identifying a transmission line structure that inherently has the desired response. The combline filter, a type of PCL filter, considered in this chapter, for example, exploits the bandpass properties of coupled microstrip transmission lines that are all shorted at the same end. Once a filter with appropriate topology has been designed, the physical realization is optimized in a circuit simulator to account for parasitics and higher-order EM effects.

While this chapter specifically addressed the design of PCL filters, the principles can be used with all distributed filter design. Bandpass filters are the most important microwave filter type and it was seen that a bandpass filter comprises coupled resonators. So other physical structures that present coupled resonators can also become components of a bandpass filter. The synthesis approach provides design insight and exploitation of all parameters of a transfer function. This leads to optimum network topologies. Synthesis can be time consuming and specialized, but it is the only way to develop filters with optimum performance.

3.7 References

[1] E. Cristal and S. Frankel, "Hairpin-line and hybrid hairpin-line/half-wave parallel-coupled-line filters," *IEEE Trans. on Microwave Theory and Techniques*, vol. 20, no. 11, pp. 719–728, Nov. 1972.

[2] B. Minnis, "A capacitively coupled bpf design using a suspended stripline," *Microwave Journal*, vol. 26, pp. 71–74, Nov. 1993.

[3] U. H. Gysel, "New theory and design for hairpin-line filters," *IEEE Trans. on Microwave Theory and Techniques*, vol. 22, no. 5, pp. 523–531, May 1974.

[4] W. Fathelbab and M. Steer, "Parallel-coupled line filters with enhanced stopband performance," *IEEE Trans. on Microwave Theory and Techniques*, vol. 53, no. 12, pp. 3774–3781, Dec. 2005.

[5] R. Wenzel, "Synthesis of combline and capacitively loaded interdigital bandpass filters of arbitrary bandwidth," *IEEE Trans. on Microwave Theory and Techniques*, vol. 19, no. 8, pp. 678–686, Aug. 1971.

[6] G. Matthaei, L. Young, and E. Jones, *Microwave Filters, Impedance-Matching Networks and Coupling Structures*. McGraw-Hill, 1965, reprinted in 1980, Artech House.

[7] J.-S. Hong and M. Lancaster, *Microstrip Filters for RF/Microwave Applications*. John Wiley & Sons, 2001.

[8] R. Mongia, I. Bahl, and P. Bhartia, *RF and Microwave Coupled-line Circuits*. Artech House, 1999.

[9] I. Hunter, *Theory and Design of Microwave Filters*. IEE Press, 2001.

[10] S. B. Cohn, "Parallel-coupled transmission-line-resonator filters," *IRE Trans. on Microwave Theory and Techniques*, vol. 6, no. 2, pp. 223–231, Apr. 1958.

[11] A. Zverev, *Foundations of Oscillator Circuit Design*. John Wiley & Sons, 1967.

[12] L. Young, *Microstrip Filters Using Parallel Coupled Lines*. Artech House, 1972.

[13] J. Rhodes, *Theory and Design of Microwave Filters*. Institute of Electrical Engineers, 2001.

[14] B. Minnis, "Printed circuit coupled-line filters for bandwidths up to and greater than an octave," *IEEE Trans. on Microwave Theory and Techniques*, vol. 29, no. 3, pp. 215–222, Mar. 1981.

[15] H. Ozaki and J. Ishii, "Synthesis of a class of strip-line filters," *IRE Trans. on Circuit Theory*, vol. 5, no. 2, pp. 104–109, Jun. 1958.

[16] B. Minnis, *Designing Microwave Circuits by Exact Synthesis*. Artech House, 1996.

[17] H. Carlin and P. Civalleri, *Wideband Circuit Design*. CRC Press, 1998.

[18] E. Cristal, "Tapped-line coupled transmission lines with applications to interdigital and combline filters," *IEEE Trans. on Microwave Theory and Techniques*, vol. 23, no. 12, pp. 1007–1012, Dec. 1975.

[19] W. Fathelbab and M. Steer, "Parallel-coupled line filters with enhanced stopband performance," *IEEE Trans. on Microwave Theory and Techniques*, vol. 53, no. 12, pp. 3774–3781, Dec. 2005.

[20] M. C. Horton and R. Wenzel, "General theory and design of optimum quarter-wave tem filters," *IEEE Trans. on Microwave Theory and Techniques*, vol. 13, no. 3, pp. 316–327, May 1965.

[21] A. Feldshtein and L. Yavich, *Synthesis of Microwave Two-Ports and Four-Ports*. Syaz, 1965.

[22] A. Riddle, "Reviewing the basics of suspended stripline," *Microwave Journal*, pp. 82–90, Oct. 2002.

[23] ——, "High performance parallel coupled microstrip filters," in *1988 IEEE MTT-S International Microwave Symposium Digest*, May 1988, pp. 427–430.

[24] L. Maloratsky, "Improve bpf performance with wiggly coupled lines," *Microwave & RF*, vol. 7, pp. 53–62, Apr. 2002.

[25] R. Rhea, *HF Filter Design and Computer Simulation*. Noble Publishing, 1994.

[26] G. Matthaei, "Comb-line band-pass filter of narrow or moderate bandwidth," *Microwave Journal*, vol. 7, pp. 428–439, Aug. 1964.

[27] R. Wenzel, "Application of exact synthesis methods to multichannel filter design," *IEEE Trans. on Microwave Theory and Techniques*, vol. 13, no. 1, pp. 5–15, Jan. 1965.

[28] W. Fathelbab and M. Steer, "Design of band-stop filters utilising circuit prototypes," *Microwaves, Antennas & Propagation, IET*, vol. 1, no. 2, pp. 523–526, 2007.

[29] M. Steer, *Microwave and RF Design, Transmission Lines*, 3rd ed. North Carolina State University, 2019.

[30] ——, *Microwave and RF Design, Networks*, 3rd ed. North Carolina State University, 2019.

[31] ——, *Microwave and RF Design, Radio Systems*, 3rd ed. North Carolina State University, 2019.

[32] I. Bahl, "Capacitively compensated high performance parallel coupled microstrip filters," in *1989 IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 1989, pp. 679–682.

[33] S.-M. Wang, C.-H. Chi, M.-Y. Hsieh, and C.-Y. Chang, "Miniaturized spurious passband suppression microstrip filter using me-

andered parallel coupled lines," *IEEE Trans. on Microwave Theory and Techniques*, vol. 53, no. 2, pp. 747–753, Feb. 2005.

[34] M. Velazquez-Ahumada, J. Martel, and F. Medina, "Parallel coupled microstrip filters with ground-plane aperture for spurious band suppression and enhanced coupling," *IEEE Trans. on Microwave Theory and Techniques*, vol. 52, no. 3, pp. 1082–1086, Mar. 2004.

[35] ——, "Parallel coupled microstrip filters with floating ground-plane conductor for spurious-band suppression," *IEEE Trans. on Microwave Theory and Techniques*, vol. 53, no. 5, pp. 1823–1828, May 2005.

[36] L. Zhu, H. Bu, and K. Wu, "Broadband and compact multi-pole microstrip bandpass filters using ground plane aperture technique," *IEE Proc., Microwaves, Antennas and Propagation*, vol. 149, no. 1, pp. 71–77, Feb. 2002.

[37] J.-T. Kuo, M. Jiang, and H.-J. Chang, "Design of parallel-coupled microstrip filters with suppression of spurious resonances using substrate suspension," *IEEE Trans. on Microwave Theory and Techniques*, vol. 52, no. 1, pp. 83–89, Jan. 2004.

[38] J.-T. Kuo and M. Jiang, "Enhanced microstrip filter design with a uniform dielectric overlay for suppressing the second harmonic response," *IEEE Microwave and Wireless Components Letters*, vol. 14, no. 9, pp. 419–421, Sep. 2004.

[39] T. Lopetegi, M. Laso, J. Hernandez, M. Bacaicoa, D. Benito, M. Garde, M. Sorolla, and M. Guglielmi, "New microstrip ldquo;wiggly-line rdquo; filters with spurious passband suppression," *IEEE Trans. on Microwave Theory and Techniques*, vol. 49, no. 9, pp. 1593–1598, Sep. 2001.

[40] T. Lopetegi, M. Laso, F. Falcone, F. Martin, J. Bonache, J. Garcia, L. Perez-Cuevas, M. Sorolla, and M. Guglielmi, "Microstrip "wiggly-line" bandpass filters with multispurious rejection," *IEEE Microwave and Wireless Components Letters*, vol. 14, no. 11, pp. 531–533, Nov. 2004.

[41] B. Kim, J. Lee, and M. Song, "An implementation of harmonic-suppression microstrip filters with periodic grooves," *IEEE Microwave and Wireless Components Letters*, vol. 14, no. 9, pp. 413–415, Sep. 2004.

[42] J.-T. Kuo, W.-H. Hsu, and W.-T. Huang, "Parallel coupled microstrip filters with suppression of harmonic response," *IEEE Microwave and Wireless Components Letters*, vol. 12, no. 10, pp. 383–385, Oct. 2002.

[43] S.-F. Chang, Y.-H. Jeng, and J.-L. Chen, "Tapped wiggly-coupled technique applied to microstrip bandpass filters for multioctave spurious suppression," *Electronics Letters*, vol. 40, no. 1, pp. 46–47, Jan. 2004.

[44] J. Garcia-Garcia, F. Martin, F. Falcone, J. Bonache, I. Gil, T. Lopetegi, M. Laso, M. Sorolla, and R. Marques, "Spurious passband suppression in microstrip coupled line band pass filters by means of split ring resonators," *IEEE Microwave and Wireless Components Letters*, vol. 14, no. 9, pp. 416–418, Sep. 2004.

[45] M. Makimoto and S. Yamashita, "Bandpass filters using parallel coupled stripline stepped impedance resonators," *IEEE Trans. on Microwave Theory and Techniques*, vol. 28, no. 12, pp. 1413–1417, Dec. 1980.

3.8 Exercises

1. A parallel coupled-line section is shown in Figure 3-11. The even-mode impedance of the coupled line is $60\ \Omega$ and the odd-mode impedance is $40\ \Omega$. The PCL section is $\lambda/4$ long at the center frequency of the section.
   (a) What is the system impedance Z_{0S}?
   (b) What is Z_{01}?
   (c) What is Z_{02}?
   (d) What is n?
   (e) What is the input impedance at Port 2 if Port 4 is terminated in $50\ \Omega$? Calculate this at the frequency at which the section is $\lambda/4$ long.

2. A parallel coupled-line section of a filter is shown in Figure 3-11. The even-mode impedance of the coupled line is $65\ \Omega$ and the odd-mode impedance is $35\ \Omega$. The PCL section is $\lambda/4$ long at the center frequency, f_0, of the section and Port 4 is terminated in $50\ \Omega$. [Hint: Consider the Z_{01} and Z_{02} lines. They are $\lambda/4$ long at f_0. Perhaps more information has been provided than you need.]
   (a) What is the input impedance at f_0 looking into Port 2?
   (b) What is the input impedance at $3f_0$ looking into Port 2?
   (c) What is the input impedance at DC looking into Port 2?

3. An interdigital coupled-line section of a filter is shown in Figure 3-12. The even-mode impedance of the coupled line is $65\ \Omega$ and the odd-mode impedance is $45\ \Omega$. The PCL section is $\lambda/4$ long at the center frequency of the filter.

(a) What is the system impedance Z_{0S}?
(b) What is Z_{01}?
(c) What is Z_{02}?
(d) What is n?
(e) What is the input impedance at Port 2 if Port 4 is terminated in 50 Ω? Calculate this at the center frequency of the filter.

4. An interdigital coupled-line section of a filter is shown in Figure 3-12. The even-mode impedance of the coupled line is 65 Ω and the odd-mode impedance is 35 Ω. The PCL section is $\lambda/4$ long at the center frequency, f_0, of the filter and Port 4 is terminated in 50 Ω. [Hint: Consider the Z_{01} and Z_{02} lines. They are $\lambda/4$ long at f_0. Perhaps more information has been provided than you need.]
   (a) What is the input impedance at f_0 looking into Port 2?
   (b) What is the input impedance at $3f_0$ looking into Port 2?
   (c) What is the input impedance at $2f_0$ looking into Port 2?

5. A combline section of a filter is shown in Figure 3-13. The even-mode impedance of the coupled line is 60 Ω and the odd-mode impedance is 40 Ω. The PCL section is $\lambda/4$ long at the center frequency of the filter.
   (a) What is the system impedance Z_{0S}?
   (b) What is Z_{01}?
   (c) What is Z_{02}?
   (d) What is n?
   (e) What is the input impedance at Port 1 if Port 2 is terminated in 50 Ω?

6. A combline section of a filter is shown in Figure 3-13. The even-mode impedance of the coupled line is 55 Ω and the odd-mode impedance is 35 Ω. The PCL section is $\lambda/4$ long at the center frequency, f_0, of the filter and Port 2 is terminated in 50 Ω. [Hint: Consider the Z_{01} and Z_{02} lines. They are $\lambda/4$ long at f_0. Perhaps more information has been provided than you need.]
   (a) What is the input impedance at f_0 looking into Port 1?
   (b) What is the input impedance at $3f_0$ looking into Port 1?
   (c) What is the input impedance at $2f_0$ looking into Port 1?

Exercises 7 to 30 cover the design of a third-order Butterworth combline filter. Most exercises begin from the solution of the previous exercise. Generally your answer for each exercise should correspond to the starting point provided for the next step in the design. You must show your detailed working.

7. Draw and derive the values of the lowpass filter prototype of a third-order Butterworth lowpass prototype filter with a corner frequency of 1 rad/s and in a 1 Ω system. Show and describe your working. [Parallels Step 1 in Section 3.4.]

8. The lowpass prototype of a third-order Butterworth filter with a corner frequency of 1 rad/s and in a 1 Ω system is shown in Figure 3-49(a), where $C_1 = 1$ F, $L_2 = 2$ H, and $C_3 = 1$ F. Transform this circuit to a lumped-element bandpass filter with corner frequencies of $f_1 = 878.42$ MHz and $f_2 = 1.1384$ GHz, and in a 50 Ω system. Draw and derive the values of the bandpass filter.

9. The lowpass prototype of a third-order Butterworth filter with a corner frequency of 1 rad/s and in a 1 Ω system is shown in Figure 3-49(a), where $C_1 = 1$ F, $L_2 = 2$ H, and $C_3 = 1$ F. Replace the series element using a shunt element and inverters. This should have electrical properties identical to those of the circuit in Figure 3-49(a). [Parallels Step 2 in Section 3.4.]

10. The lowpass prototype of a third-order Butterworth filter with a corner frequency of 1 rad/s and in a 1 Ω system is shown in Figure 3-49(a), where $C_1 = 1$ F, $L_2 = 2$ H, and $C_3 = 1$ F. Replace the series element using a shunt element and inverter(s) and draw and derive the values of the new prototype now with a corner frequency of 1 Hz. This should have electrical properties identical to those of the circuit in Figure 3-49(a). [Parallels Step 2 in Section 3.4.]

11. Consider eliminating the transformer in Figure 3-49(b).
   (a) What effect does this have on the properties of the lowpass filter?
   (a) Draw and derive the values of the new prototype without the transformer.

12. Derive the bandpass filter prototype of a third-order bandpass filter with corner frequencies of $f_1 = 878.42$ MHz and $f_2 = 1.1384$ GHz. Base this on the 1 Ω third-order Butterworth lowpass prototype shown in Figure 3-49(c) with $C_1 = C_3 = 1$ F and $C_2 = 2$ F.
   (a) What is the center frequency of the bandpass filter?
   (b) What is the fractional bandwidth?
   (c) First consider a 1 Ω system (i.e., the source and load resistances are 1 Ω). Use two 1 Ω impedance inverters, but use no stubs. Draw and derive the values of the new prototype. [Parallels Step 3 in Section 3.4.]
   (d) Transform the prototype developed in (a) to a 50 Ω system (i.e., the source and load resis-

tances are 50 Ω). Draw and derive the values of the new prototype. [Parallels Step 4 in Section 3.4.]

13. Consider a shorted stub that is resonant at frequency f_r.
    (a) What is the input impedance of the stub at resonance (this should be taken as the first resonance)?
    (b) In terms of wavelengths, how long is the stub at f_r?
    (c) In terms of wavelengths, how long is the stub at $\frac{1}{2}f_r$?

14. Consider an open-circuited stub that is resonant at frequency f_r.
    (a) What is the input impedance of the stub at resonance (this should be taken as the first resonance)?
    (b) In terms of wavelengths, how long is the stub at f_r?
    (c) In terms of wavelengths, how long is the stub at $\frac{1}{2}f_r$?

15. Consider a capacitor C.
    (a) What is the admittance with respect to radian frequency of the capacitor at frequency $f_0 = \omega_0/2\pi$?
    (b) What is the derivative of the admittance of the capacitor with respect to radian frequency of the capacitor of the resonator at f_0?

16. A resonator comprising a capacitor C in parallel with an inductor L is resonant at a frequency f_0.
    (a) What is the admittance of the resonator at f_0?
    (b) What is the derivative with respect to radian frequency of the admittance of the resonator at f_0?

17. Consider a shorted stub that is resonant at frequency $f_r = 2f_0$.
    (a) In terms of wavelengths, how long is the stub at f_r?
    (b) In terms of wavelengths, how long is the stub at f_0?
    (c) What is the admittance of the stub at f_0?
    (d) What is the derivative with respect to radian frequency of the admittance of the stub at f_0?

18. A lumped-element resonator consists of a parallel capacitor $C_1^\sharp = 12.2427$ pF and inductor $L_1^\sharp = 2.06901$ nH. Derive an equivalent resonator comprising a capacitor in parallel with a shorted stub that is resonant at the frequency $f_r = 2f_0$. The new resonator must have the same admittance at $f_0 = 1$ GHz as the original resonator and the derivative of the admittance must be the same at f_0 for both resonators. Draw and derive the values of the new resonator. [Parallels Step 5 in Section 3.4.]

19. The prototype of a 50 Ω third-order Butterworth bandpass filter is shown in Figure 3-49(e). The filter has a center frequency of $f_0 = 1$ GHz, where $C_1^\sharp = C_3^\sharp = 12.2427$ pF, $L_1^\sharp = L_3^\sharp = 2.06901$ nH, $C_2^\sharp = 24.4854$ pF, and $L_2^\sharp = 1.03451$ nH. Replace each of the three lumped-element resonators by a resonator consisting of a single capacitor and a shorted stub. Each of the new resonators must have the same admittance at f_0 as the original resonator it replaced and the derivative with respect to frequency of the admittances of each of the original and replacement resonators must be the same at f_0. Draw and derive the values of the new filter prototype. [Parallels Step 5 in Section 3.4.]

20. A shorted stub has a characteristic impedance of $Z_{01} = 75$ Ω and is resonant at the frequency $f_r = 2f_0$.
    (a) What is the input impedance of the stub at f_r?
    (b) What is the input impedance of the stub at f_0?

21. The prototype of a 50 Ω third-order Butterworth bandpass filter is shown in Figure 3-49(f). The filter has a center frequency of $f_0 = 1$ GHz, where $C_1'' = C_3'' = 9.52443$ pF, $Z_{01} = Z_{03} = 16.7102$ Ω, $C_2'' = 19.0489$ pF, $Z_{02} = 8.35509$ Ω. The resonant frequency of each of the stubs is $f_r = 2f_0$. [Parallels Step 6 in Section 3.4.]
    (a) What is the input impedance at f_0 of the stub with characteristic impedance Z_{01}?
    (b) Transform the prototype so that each stub has the characteristic impedance Z_{01}.

22. An impedance inverter has a characteristic impedance of 50 Ω. Develop the lumped-element equivalent circuit of the inverter. The equivalent circuit should have three lumped impedances. Draw and derive the values of the equivalent circuit.

23. An impedance inverter has a characteristic impedance of 60 Ω. Develop the lumped-element equivalent circuit of the inverter at 1 GHz. The equivalent circuit should have three lumped elements. Draw and derive the values of the equivalent circuit with inductor and capacitor values.

24. The prototype of a 50 Ω third-order Butterworth bandpass filter is shown in Figure 3-49(g). The filter has a center frequency of $f_0 = 1$ GHz,

where $C_1'' = C_2''' = C_3'' = 9.52443$ pF, $Z_{01} = Z_{02}' = Z_{03} = 16.7102\ \Omega$, and $Z_y = 70.7107\ \Omega$. The resonant frequency of the stubs is $f_r = 2f_0$. Draw and derive the values of the filter prototype with the inverters replaced by short-circuited stubs resonant at f_r. [Parallels Step 7 in Section 3.4.]

25. The prototype of a $50\ \Omega$ third-order Butterworth bandpass filter is shown in Figure 3-50(h). The filter has a center frequency of $f_0 = 1$ GHz, where $C_1'' = C_3'' = 9.52443$ pF, $Z_{01}' = Z_{03}' = 21.8811\ \Omega$, $C_2''' = 9.52443$ pF, $Z_{02}'' = 31.6862\ \Omega$, and $Z_{012} = Z_{023} = 70.7107\ \Omega$. The resonant frequency of the stubs is $f_r = 2f_0$. The characteristic impedances Z_{01}', Z_{02}'', and Z_{03}' are too low to be realized in microstrip and so need to be scaled to a more reasonable microstrip impedance (say between $30\ \Omega$ and $80\ \Omega$). Scale the middle stub, Z_{02}'' to $80\ \Omega$. This will shift the source and load impedances away from $50\ \Omega$, but this can be accommodated in a latter step. [Parallels Step 8 in Section 3.4.]

26. The prototype of a third-order Butterworth bandpass filter is shown in Figure 3-50(i). The filter has a center frequency of $f_0 = 1$ GHz, where $R_S = R_L = 126.238\ \Omega$, $C_{t1} = C_{t2} = C_{t3} = 3.77241$ pF, $Z_{0t1} = Z_{0t3} = 55.2444\ \Omega$, $Z_{0t2} = 80\ \Omega$, and $Z_{0t12} = Z_{0t23} = 178.528\ \Omega$. The resonant frequency of the stubs is $f_r = 2f_0$. Incorporate an inverter at the input and at the output of the filter so that it interfaces with $50\ \Omega$ source and load impedances. Draw and derive the values of the new prototype. [Parallels Step 9 in Section 3.4.]

27. The prototype of a third-order Butterworth bandpass filter is shown in Figure 3-50(j). The filter has a center frequency of $f_0 = 1$ GHz, where the impedance inverters have impedances $Z_1 = Z_2 = 79.4475\ \Omega$, $C_{t1} = C_{t2} = C_{t3} = 3.77241$ pF, $Z_{0t2}'' = 80\ \Omega$, $Z_{0t1} = Z_{0t3} = 55.244\ \Omega$, and $Z_{0t12} = Z_{0t23} = 178.5275\ \Omega$. The resonant frequency of the stubs is $f_r = 2f_0$. Incorporate an inverter at the input and and at the output of the filter so that it interfaces with $50\ \Omega$ source and load impedances. Draw and derive the values of the new prototype. Realize that each of the impedance inverters using a two-capacitor network noting that one side of the inverters is a resistance. Note that one of the capacitor values in each network will be negative. [Parallels Step 9 in Section 3.4.]

28. The prototype of a third-order Butterworth bandpass filter is shown in Figure 3-50(k). The filter has a center frequency of $f_0 = 1$ GHz where $C_b = 2.57780$ pF, $C_1 = C_3 = 2.21562$ pF, $Z_{0t1} = Z_{0t3} = 55.2444\ \Omega$, $Z_{0t2}'' = 80\ \Omega$, $Z_{0t12} = Z_{0t23} = 178.5275\ \Omega$, $C_{t2} = 3.77241$ pF. Sketch the physical layout of the circuit assuming that lumped-element capacitors will be used. You need not develop the dimensions of the microstrip lines. [Parallels Step 9 in Section 3.4.]

29. Design a third-order maximally flat bandpass filter prototype in a $50\ \Omega$ system centered at 1 GHz with a 10% bandwidth. The lowpass prototype of a third-order maximally flat filter is shown in Figure 2-10. The problem will parallel the development in Section 3.4 and the end result of this development will be a bandpass prototype filter with the form of that in Figure 3-29. Note that there will be differences as the filter is a different type.
    (a) Convert the prototype lowpass filter to a lowpass filter with inverters and capacitors only; that is, remove the series inductors.
    (b) Scale the filter to take the corner frequency from 1 rad/s to 1 GHz.
    (c) Transform the lowpass filter into a bandpass filter. That is, replace each shunt capacitor by a parallel LC network. This step will establish the bandwidth of the filter.
    (d) Transform the system impedance of the filter from 1 to $50\ \Omega$.
    (e) Replace the parallel LC circuits by short-circuited stubs in parallel with lumped capacitors. (The circuit will now be in a form similar to that in Figure 3-25.)
    (f) For each inverter, derive the three-lumped-element equivalent circuit as in Figure 2-20. Do not update the filter prototype yet, but instead draw and label the lumped-element equivalent circuits of each inverter.

30. A two-port consisting of a three shorted stubs in a Pi structure with the shunt stubs having a characteristic impedance of $55.2444\ \Omega$ and the series stub has a characteristic impedance of $178.5275\ \Omega$. The operating frequency is $f_0 = 1$ GHz and the resonant frequency of the stubs is $f_r = 2f_0$. The substrate has a thickness of $500\ \mu$m and a relative permittivity ε_r of 10.
    (a) Draw the inverter.
    (b) Draw the pair of coupled lines in combline configuration and draw two equivalent circuit models of the combline. Note that the coupled lines are a $\lambda/8$ long at f_0. Calculate the element values in your equivalent models.
    (c) Calculate two values for the system impedance. Take the system impedance as the geometric mean of these two values.

(d) Calculate the odd-mode impedance of the coupled lines. (Use $50\ \Omega$ if you were not able to solve Part (c).)
(e) Calculate the even-mode impedance of the coupled lines. (Use $50\ \Omega$ if you were not able to solve Part (c).)
(f) Calculate the width and separation of the coupled lines. Use may want to use Figures 5-10 to 5-13 and/or Tables 5-2 and 5-3 of [29]. Also, you may need to interpolate. Use the tables or figures even if your system impedance is not $50\ \Omega$.
(g) What is the effective permittivity of the even mode?
(h) What is the effective permittivity of the odd mode?
(i) Determine the length of the coupled lines taking the average effective permittivity as the geometric mean of the even- and odd-mode values.

31. The prototype of a third-order Butterworth bandpass filter is shown in Figure 3-50(k). The filter has a center frequency of $f_0 = 1$ GHz, where $C_b = 2.57780$ pF, $C_1 = C_3 = 2.21562$ pF, $Z_{0t1} = Z_{0t3} = 55.2444\ \Omega$, $Z''_{0t2} = 80\ \Omega$, $Z_{0t12} = Z_{0t23} = 178.5275\ \Omega$, $C_{t2} = 3.77241$ pF. Sketch the physical layout of the circuit assuming that lumped-element capacitors will be used. Calculate the widths and lengths of the microstrip lines if the thickness of the microstrip substrate is $500\ \mu$m and the relative permittivity of the substrate is 10. Use can use, with some error, Figures 5-10–5-13 and/or Tables 5-2–5-3 of [29] in calculating the widths and separations of the coupled lines even if your system impedance is not $50\ \Omega$.

3.8.1 Exercises by Section

†challenging, ‡very challenging

§3.2 $1^{\dagger}, 2^{\dagger}, 3^{\dagger}, 4^{\dagger}, 5^{\dagger}, 6^{\dagger}$
§3.4 $7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18^{\dagger}, 19^{\dagger}, 20, 21^{\dagger}, 22, 23, 24^{\dagger}, 25^{\dagger}, 26^{\dagger}, 27^{\dagger}, 28^{\dagger}, 29^{\ddagger}, 30^{\ddagger}, 31^{\ddagger}$

3.8.2 Answers to Selected Exercises

1(d) $2\ \Omega$
3 5.5
5(d) 5
6(c) $0\ \Omega$
14 $\lambda/8$
16(b) $j[C+1/(\omega_0^2 L)]$
18 9.5 pF
20(a) ∞
30(d) 1.48 cm
31 $39.4\ \Omega$

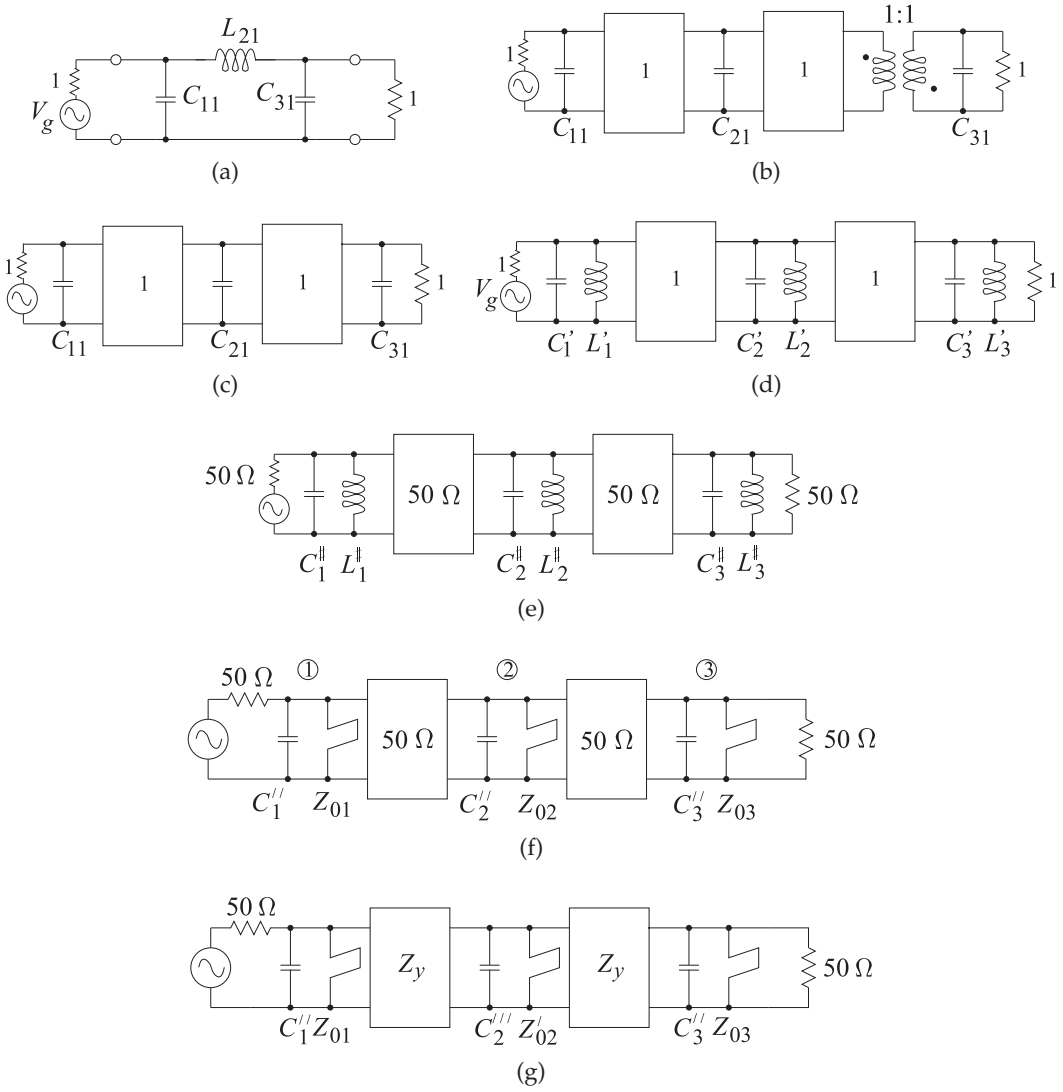


Figure 3-49: Early prototypes in the development of a third-order Butterworth combline filter.

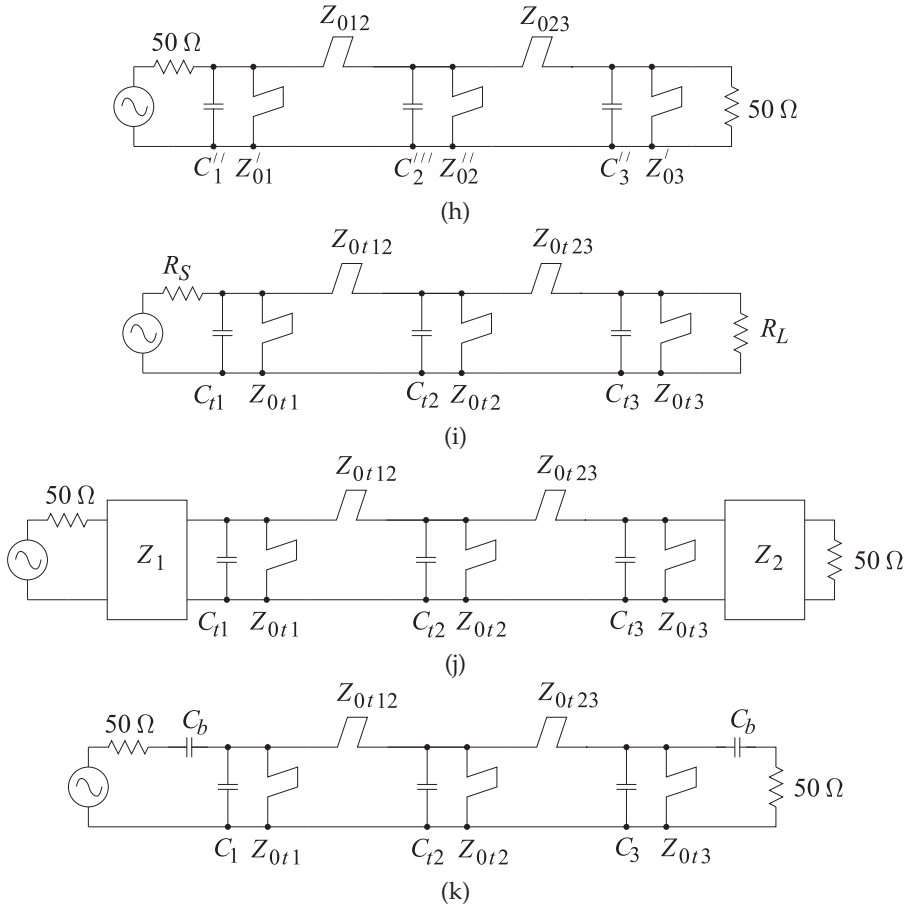


Figure 3-50: Late-stage prototypes in the development of a third-order Butterworth combline filter.

CHAPTER 4

Noise, Distortion, and Dynamic Range

4.1 Introduction 123
4.2 Noise 123
4.3 Noise Characterization 130
4.4 Oscillator Noise 141
4.5 Nonlinear Distortion 144
4.6 Dynamic Range 151
4.7 Passive Intermodulation Distortion 154
4.8 Breakdown 156
4.9 Summary 158
4.10 References 159
4.11 Exercises 161

4.1 Introduction

Measures of noise, distortion, and dynamic range are system metrics describing the sensitivity of active systems and modules. In this chapter the limitations described by these metrics will be discussed in a form that is applicable to both active device design and system design using modules. Noise and nonlinear distortion set the bounds on the range of signals that can be processed by an RF circuit. Noise (i.e., random fluctuations of voltage and current) establishes the minimum detectable signal, while nonlinear distortion sets the level of the largest signal from which information can be reliably extracted or which can be transmitted without affecting other systems.

4.2 Noise

Fluctuations of voltage and current arise from several different physical processes yielding noise with various statistical properties. Various types of noise are important in electronic circuits and these types range from noise, such as thermal noise, that is very well understood, to noise that has been observed and seriously affect the performance of RF circuits but not well understood. An example of the latter is phase noise on oscillators which manifests itself as random fluctuations of the phase or frequency of an oscillation signal. One of the problems in understanding noise is that it can be difficult to describe a particular type of noise in both the time and frequency

domains. Not all types of noise can be described in a straightforward way and the sources of some types of noise are not understood. The impact of noise on an RF system is described in the frequency domain whereas the physical origins of noise must necessarily be in the real world (i.e., the time domain). The impact of noise is also in the time domain resulting in bit errors of a received, demodulated, and processed digitally modulated signal. Noise originating from a noise source is shaped by the characteristics of a circuit before it is observed externally and thus the true nature of noise is further obscured.

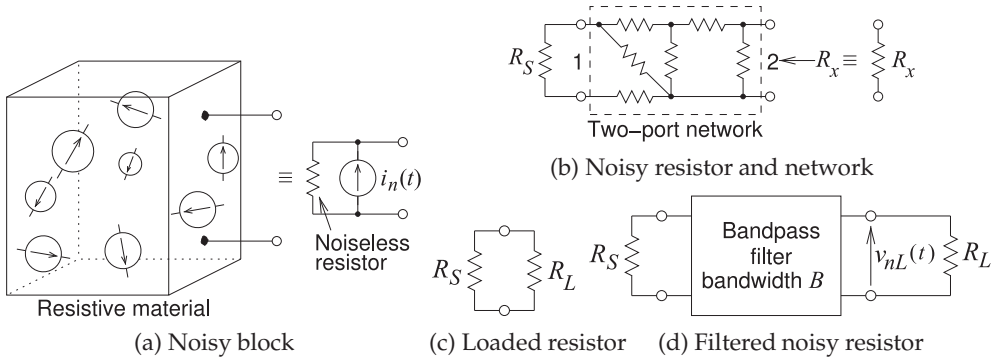
4.2.1 Observations of Noise Spectra

Noise in electronics is attributed to the random movement of carriers and most types of noise have a power spectral density that is flat with respect to frequency. Such noise is called white noise and if it is filtered, lowpass filtered or bandpass filtered, it is called additive white Gaussian noise (AWGN) as the noise then has an asumed Gaussian statistical distribution. The best understood noise is thermal noise and is attributed to the random movement of electrons due to the random vibration of the lattice of a conducting material. The theory of thermal noise is based on the fluctuation-dissipation theorem [1] which can be used for most materials in thermal equilibrium. This theorem applies both to classical and quantum mechanical systems and describes the classical noise encountered up to several terahertz as well as the quantum-mechanical effects that shape noise above a few terahertz at room temperature or at much lower frequencies as temperature approaches absolute zero [2].

The fluctuation-dissipation theorem relates thermally-induced fluctuations in a material to the resistance of the material. The physical origin of thermal noise is the net effect of the rapidly-fluctuating currents resulting from thermal fluctuations of free electrons in a resistive (or conductive) material. The fluctuations of the electrons result from the vibrations of the atoms in a material's lattice. These vibrations relate directly to temperature and temperature is regarded as a direct measure of the state of entropy or random vibration of the lattice. So the fluctuation-dissipation theorem describes how vibration (i.e. temperature) of the lattice induces small current fluctuations in a material and thus describes how heat energy is converted into the electrical energy known as Johnson noise. This is the opposite of the effect of resistance which converts electrical energy into heat energy when the movement of electrons (as current) causes the lattice to vibrate more. So it is not surprising that the noise current is directly related to resistance.

Key results of the fluctuation-dissipation theorem are that the available noise power from a resistor or group of resistors is linearly proportional to temperature, the available noise power is independent of the resistor value, and that the power spectral density (in watts per hertz) is independent of frequency, i.e. it is white, up to a few terahertz at room temperature.

Consider the block of resistive material in Figure 4-1(a). Lattice vibrations cause random movement of electrons and thus there is a myriad of tiny current sources. The block of resistive material has terminals and at those terminals a resistance can be measured. Application of the fluctuation-dissipation theorem determines that the net effect of the little noise-current sources in the material is the same as that of a current source in parallel



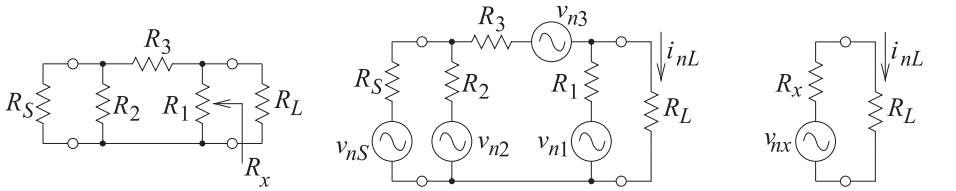
(a) Noisy block (b) Noisy resistor and network (c) Loaded resistor (d) Filtered noisy resistor

Figure 4-1: Noisy resistive networks

with the effective resistance as shown in Figure 4-1(a). The noise current source, $i_n(t)$, has statistical properties such that the available noise power is proportional to the bandwidth over which the noise power is measured and to the temperature of the material in kelvin. If the resistive material is at a temperature T_0, then the noise temperature of the material T_0. Going one step further, it can be imagined that the resistive material can be described by a network of resistors as shown in Figure 4-1(b). If R_S and each resistor in the network is at the same temperature T_0, then the whole network, including R_S, is equivalent both in terms of resistance and noise to a single noisy resistance R_x, that is at T_0. So the noise temperature of the resistive network is equal to that of R_x both being T_0. This is true provided that there are no non-thermal sources of noise inside the two-port network. For example, if there were transistors then an additional source of noise is shot noise. Then the noise temperature looking into port 2 of the network would be greater than the noise temperature of the source resistance at port 1.

In Figure 4-1(c) a noisy resistance R_S has an available noise power and all of that available noise power is delivered to the load R_L provided that $R_L = R_S$. (All of the available noise power from R_L will also be delivered to R_S under matched conditions.) Even if a resistor is not loaded, so R_S is on its own, there will be noise power in the resistor which will be constantly generated and reabsorbed (through resistive heating) so that the resistor is in thermal equilibrium.

A noisy resistor is equivalent both electrically and from a noise perspective to a noise-free resistor with a shunt noise current source as in Figure 4-1(a) or equivalently as the same value of resistance with a noise voltage source (which will be introduced latter). The noise voltage and current sources are random and if the noise is lowpass or bandpass filtered as in Figure 4-1(d), the resulting noise voltage $v_{nL}(t)$ across the noise-free load will have Gaussian statistics. Since it is inevitable that noise will be filtered in a circuit, e.g. there will be at least parasitic capacitances, thermal noise is often treated as being additive white Gaussian noise (AWGN) as its statistical properties will be Gaussian. This is indeed fortuitous as it is possible to greatly simplify the treatment of noise if it can be consider to be random with Gaussian statistics. This is exploited in the development of the mathematics of random processes in Appendix 1.A of [3] as it applies to both noise and



(a) Noisy resistor network (b) Noise-free resistor network (c) Noise-free resistor

Figure 4-2: Attenuator example.

digitally modulated signals. The key result is that noise can be described in the frequency domain and this understanding and characterization can be translated to the real world, i.e. the time domain.

As an example, it is possible to undertake a frequency domain analysis of the circuit shown in Figure 4-2(a) where a source R_S is connected to a resistive attenuator and then to a noise-free load R_L. As usual the matched resistances external to the two-port are the same here so that $R_S = R_L$. The network of Figure 4-2(a) can be replaced by a network of noise-free resistors with noise voltage sources v_{nS}, v_{n1}, v_{n2}, and v_{n3}. The noise voltage sources are random and independent and so are uncorrelated. Thus to evaluate the noise current i_{nL} in R_L the powers of the individual contributions to i_{nL} need to be summed first. This can be shown to be identical to calculating i_{nL} in Figure 4-2(c) where the equivalent resistance R_x of the network is found and then the noise voltage source, v_{nx} for that resistor used. Of course since the two-port is an attenuator $R_x = R_S = R_L$.

4.2.2 Characterization of Thermal Noise

While the impact of noise on RF circuits is measured and categorized in the frequency domain, the physical sources of noise are in the real world. Noise in a conductor is manifested as random fluctuations in time of voltage and current. While random, the noise can have different statistics depending on how it originates. The three major physical sources of noise affecting electronic circuits are thermal, shot, and flicker.

Thermal noise is more formally known as Johnson-Nyquist noise and is also referred to as Johnson noise or Nyquist noise. The noise is due to random fluctuations of charge carriers inside a conductor occurring with or without applied voltage or current. A resistor at room temperature (290–298 K or 19–25°C) has an available noise power of -174 dBm in 1 hertz of bandwidth. The noise is uncorrelated so that the noise power in a second hertz of bandwidth will add. So in 2 hertz of bandwidth the available noise power is $(-174 + 3)$ dBm $= -171$ dBm.

The extent of fluctuation is linearly proportional to absolute temperature. Also the noise power generated is independent of the resistance of the conductor. The original derivation of thermal noise is due to Nyquist [4], who showed that the power spectral density (PSD) of the available noise power from a resistor (of any value) is

$$S_t(f) = kT, \tag{4.1}$$

where k is the Boltzmann constant and T is the temperature in kelvin. The subscript t here is used to indicate thermal noise. The SI units of S_t are watts per hertz but is more commonly expressed as dBm per hertz. The available noise power in a bandwidth B is (in units of watts)

$$P_t(f) = kTB. \tag{4.2}$$

When quantum effects are important, Equation (4.1) is modified and the thermal noise PSD is

$$S_t(f) = \frac{hf}{e^{hf/kT} - 1}, \tag{4.3}$$

where h is Plank's constant. In Equation (4.3) hf is the energy of a photon of frequency f and kT is the average thermal energy (i.e. vibrational kinetic energy) of the material. To determine the frequency at which the simpler form of S_t can be used consider the following. At low to moderate frequencies Equation (4.3) can be expanded as

$$S_t(f) = \frac{hf}{1 + (hf/kT) + \frac{1}{2}(hf/kT)^2 + \ldots - 1} \approx \frac{kT}{1 + \frac{1}{2}(hf/kT)}. \tag{4.4}$$

Thus the thermal noise power available will drop off as frequency increases, and is at one-half its low-frequency value at a critical frequency $f_c = 2kT/h$. At room temperature this is approximately 12 THz. So quantum effects on thermal noise are not of concern at room temperature at frequencies below a few terahertz.

Extensive modern treatments of thermal noise are available in [5] and [6].

EXAMPLE 4.1 Available Noise Power

What is the available noise power from a resistor in a 50 MHz bandwidth and at 20°C.

Solution:

The PSD is, from Equation (4.1),

$$\begin{aligned} S_t(f) &= (1.381 \cdot 10^{-23}\ \text{J/K}) \cdot ((273 + 20)\ \text{K}) \\ &= 4.046 \cdot 10^{-21}\ \text{J} = 4.046\ \text{zJ} = 4.046\ \text{zW/Hz} = -173.9\ \text{dBm/Hz}. \end{aligned} \tag{4.5}$$

That is, 4.046 zJ (zepto joules). The thermal noise at room temperature is usually taken as -174 dBm/Hz. This noise power is equally divided between amplitude noise and phase noise (each is -177 dBm/Hz) [7]. $S_t(f)$ is multiplied by the bandwidth to obtain the total available thermal noise power, so for a 50 MHz bandwidth, the thermal noise power is

$$\begin{aligned} P_t(f) &= (4.046 \cdot 10^{-21}\ \text{J}) \times (50 \cdot 10^6\ \text{Hz}) \\ &= 2.023 \cdot 10^{-13}\ \text{J} \cdot \text{s}^{-1} = 202.3\ \text{fW} = -100.1\ \text{dBm}. \end{aligned} \tag{4.6}$$

This is an appreciable power given that cell phones can operate with receive signals smaller than -90 dBm. So the lesson here is to use the smallest bandwidth possible in designs.

4.2.3 Environmental Noise

Noise in RF and microwave systems includes noise from the environment as well as noise generated within the circuitry itself. Noise from the environment can have galactic origins, when it is known as cosmic background noise, from black-body radiation, or can be artificially generated noise. In cellular communication systems the major source of interference is from other phones and base stations in the cellular system. Provided this is uniformly random over the communication band, it can be treated as random noise. Uniformly random noise (i.e., white noise) can be modeled by a resistor held at what is called the noise temperature.

A noisy resistor generates **white noise** that has a flat PSD, i.e. noise spectral density is independent of frequency. A noisy resistor can be modeled by a noise-free resistor and a random voltage or current source denoted by v_n and i_n, respectively (see Figure 4-3). The sources v_n and i_n are random and their spectral densities, S_{vn} and S_{in} respectively, are related by

$$S_{vn}(f) = R^2 S_{in}(f). \tag{4.7}$$

The noise voltage spectral density of a resistor R at temperature T is [8]

$$S_{vn}(f) = \frac{v_n^2}{B} = 4kTR, \tag{4.8}$$

so that

$$v_n^2 = 4kTBR. \tag{4.9}$$

Here k is the Boltzmann constant, T is the temperature in kelvins, and B is the bandwidth in hertz.

It is possible for the noise temperature looking into a two-port to be less than the ambient temperature and then the effective noise temperature of the structure is used as a measure of available noise power. A typical situation is specifying the noise presented by an antenna to a receiver. The noise captured by an antenna is from the environment with usually only a small portion of the noise coming from the antenna itself. In the absence of antenna loss, an antenna pointed into space will have a noise temperature corresponding to the cosmic microwave background radiation with an effective noise temperature of about 3 K.

Two noise voltage sources, v_{n1} and v_{n2}, in series can be partly correlated. Then the two noise sources can be replaced by a single source v_n, where

$$v_n^2 = v_{n1}^2 + v_{n2}^2 + 2C_{n1,n2}v_{n1}v_{n2}. \tag{4.10}$$

Here $C_{n1,n2}$ is the **correlation coefficient**, and $-1 \leq C_{n1,n2} \leq 1$. If the sources are uncorrelated, as they would be for two resistors, $C_{n1,n2} = 0$ and

$$v_n^2 = v_{n1}^2 + v_{n2}^2. \tag{4.11}$$

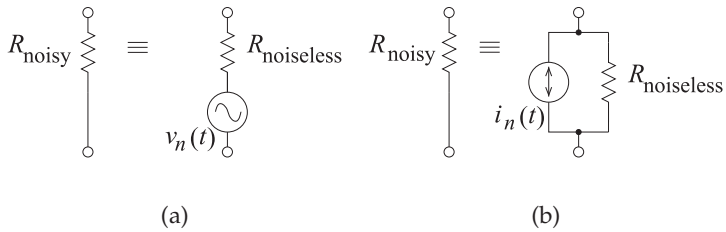


Figure 4-3: Thermal noise equivalent circuits: (a) noisy resistor modeled as a noise-free resistor in series with a random noise voltage source v_n; and (b) noisy resistor modeled as a noise free resistor in parallel with a random noise current source i_n.

Correlation of noise sources is important to modeling noise in transistors, as there can be a common physical origin for noise that is modeled as two noise sources in a circuit model.

4.2.4 Thermal Noise and Capacitors

Thermal noise is also seen with reactive components such as a capacitor where it is known as kT/C (read as k-T-C) noise. The series resistance, R, of a capacitor, C, contributes the thermal noise but the RC combination also filters the noise. This noise can be treated the same way as the thermal noise analysis above, but there is a short-hand way of looking at the noise. Sarpeshkar et al. [9] showed that the mean-square noise voltage on a capacitor of value C within the noise bandwidth of the RC circuit (derived as $1/(4RC)$ in hertz) is

$$\bar{v}_n^2 = \frac{kT}{C}. \tag{4.12}$$

At room temperature (i.e., $T = 20°\mathrm{C} = 293\ K$), the root mean square noise voltage, $\sqrt{\bar{v}_n^2}$, on a 10 pF capacitor with a $1\ \Omega$ series resistance is $20\ \mu\mathrm{V}$ in a 25 GHz bandwidth.

4.2.5 Physical Source of Shot Noise

Shot noise is due to current being carried by discrete charge carriers. It is important when there is a region that is scarce of free carriers. Shot noise is particularly important with semiconductor devices but was first observed by Schottky in 1926 in vacuum tubes. In a semiconductor the charge carriers, electrons and holes, are discrete and independent. As such the current fluctuates as the number of carriers varies in discrete steps. On average there is a net velocity of carriers passing a point per time interval. For shot noise to be observed above thermal noise, the carriers should be constrained to pass in just one direction. This is the situation in many semiconductor devices where the depletion region formed at the interface of pn junctions forces the current to flow in just one direction. Shot noise is more significant when the number of charge carriers is small, a situation that also exists in semiconductors. However, even in a semiconductor, there are enough carriers for shot noise to have a Gaussian distribution so that statistically it looks like thermal noise [10]. The shorter the critical time scale, and for microwave and RF circuits this is the period of the waveform, the fewer the number of carriers that will pass a point and the greater the fractional contributions of fluctuations in carrier numbers.

The RMS current fluctuation due to shot noise is

$$\sigma_i = \sqrt{2eIB}, \tag{4.13}$$

where e is the elementary charge, B is the bandwidth in hertz, and I is the current. For a DC current of 1 mA and in a 1 Hz bandwidth, the RMS current fluctuation due to shot noise is $\sigma_i = 18$ pA. If this current flows through a resistance R, the spectral noise density in the resistor is

$$P_s = 2e|I|. \tag{4.14}$$

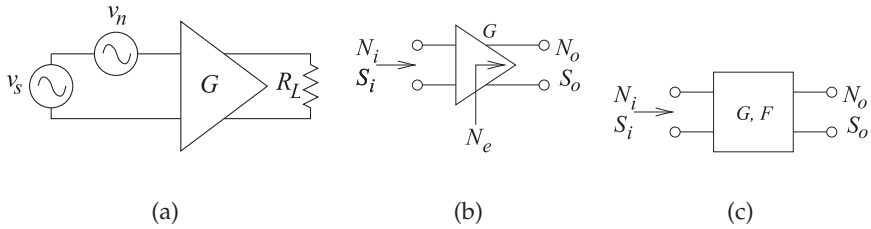


Figure 4-4: Noise and two-ports: (a) amplifier; (b) amplifier with excess noise; and (c) noisy two-port network.

Note that P_s is independent of temperature and frequency, although eventually quantum effects become important [11]. In contrast, thermal noise is proportional to temperature. The current in Equation (4.14) is the instantaneous current. So shot noise varies during an RF cycle as the current flow varies. If the DC current is much larger than the time-varying current, then the noise that is created by the separable pulses has a flat frequency spectrum and can be modeled by a white noise source.

So the lesson here is that in active circuit designs, bias currents should be minimized. Also note that the noise power is a function of the resistance value but active circuits with high resistance values also tend to have low current levels.

4.2.6 Physical Source of Flicker Noise

The third type of noise that is of concern with RF and microwave circuits is flicker noise, sometimes called $1/f$ (one-over-f) noise because of its power spectral density shape. Flicker noise is due to diffusion, traps in a semiconductor, and surface traps. A free carrier is immobilized or trapped when it falls into a trap, that is, a recombination center. When several such carriers are trapped, it means that they are not available for conduction and as a result, the resistance of the semiconductor is modulated. These fluctuations have multiple relaxation times. Flicker noise is considered again in Section 6.4 in regards to the characterization of local oscillator modules. A complete physical understanding of flicker noise is not available, and flicker noise is a major concern with oscillators.

4.3 Noise Characterization

Amplifiers, filters, and mixers in an RF front end process (e.g., amplify, filter, and mix) input noise the same way as an input signal. In addition, these components contribute **excess noise** of their own. Without loss of generality, the following discussion considers noise with respect to the amplifier shown in Figure 4-4(a), where v_s is the input signal. The development here applies only to white noise (i.e., thermal and shot noise). The noise signal, with source designated by v_n, is uncorrelated and random, and described as an RMS voltage or by its noise power.

The most important noise-related metric is the SNR. Denoting the noise power input to the amplifier as N_i, and denoting the signal power input to the amplifier as S_i, the input signal-to-noise power ratio is $\text{SNR}_i = S_i/N_i$. If the amplifier is noise free, then the input noise and signal powers are amplified by the power gain of the amplifier, G. Thus the output noise power is $N_o = GN_i$, the output signal power is $S_o = GS_i$, and the output SNR is $\text{SNR}_o = S_o/N_o = \text{SNR}_i$.

In practice, an amplifier is noisy, with the addition of excess noise,

N_e, indicated in Figure 4-4(b). The excess noise originates in different components in the amplifier and is either referenced to the input or to the output of the amplifier. Most commonly it is referenced to the output so that the total output noise power is $N_o = GN_i + N_e$. In the absence of a qualifier, the **excess noise** should be assumed to be referred to the output. N_e is not measured directly. Instead, the ratio of the SNR at the input to that at the output is the **noise factor**:

$$F = \frac{\text{SNR}_i}{\text{SNR}_o}, \tag{4.15}$$

and this is the way noise performance is normally measured. If the circuit is noise free, then SNR_o = SNR_i and $F = 1$. If the circuit is not noise free, then $\text{SNR}_o < \text{SNR}_i$ and $F > 1$. F can be related to the excess noise produced in the circuit. With the excess noise, N_e, referred to the circuit output,

$$\begin{aligned} F &= \frac{\text{SNR}_i}{\text{SNR}_o} = \frac{\text{SNR}_i}{1}\frac{1}{\text{SNR}_o} = \frac{S_i}{N_i}\frac{N_o}{S_o} = \frac{S_i}{N_i}\frac{GN_i + N_e}{GS_i} \\ &= 1 + \frac{N_e}{GN_i}. \end{aligned} \tag{4.16}$$

One of the conclusions that can be drawn from this is that the noise factor, F, depends on the available noise power at the input of the circuit. As a standard reference, the available noise power, N_R, from a resistor at **standard temperature**, T_0 (290 K) [12], and over a bandwidth, B (in Hz), is used,

$$N_i = N_R = kT_0B, \tag{4.17}$$

where k ($= 1.381 \times 10^{-23}$ J/K) is **the Boltzmann constant**. If the input of an amplifier is connected to this resistor and all of the noise power is delivered to the amplifier, then

$$F = 1 + \frac{N_e}{GN_i} = 1 + \frac{N_e}{GkT_0B}. \tag{4.18}$$

Several random physical processes inside a circuit contribute to excess noise, and not all of these processes vary linearly with temperature. Consequently F is a function of temperature, although usually a weak one. It is also a function of bandwidth, and there is a problem in using F with cascaded systems in which bandwidths vary for different subsystems. Even with all these problems, F is the most important measure used to characterize noise performance. It can be used to determine the noise performance of a cascade, when the noise factors and gains of the subsystem constituents are known. F is the ratio of powers, and when expressed in decibels, the **noise figure** (**NF**) is used:

$$\begin{aligned} \text{NF} &= 10\log_{10} F \\ &= \text{SNR}_i|_{\text{dB}} - \text{SNR}_o|_{\text{dB}}\,, \end{aligned} \tag{4.19}$$

where the SNR is expressed in decibels.

Consider the amplifier in Figure 4-4. If the excess noise contribution of an amplifier is ignored, the output noise power will be

$$N_o = GkT_0B. \tag{4.20}$$

With the amplifier's excess noise, N_e, included, the output noise power is

$$\begin{aligned} N_o &= GkT_0B + N_e = GkT_0B(1 + N_e/(GkT_0B)) \\ &= FGkT_0B. \end{aligned} \tag{4.21}$$

Rearranging this equation, the excess noise power can be written as

$$N_e = (F-1)GkT_0B. \tag{4.22}$$

The output noise of a system can be expressed in terms of its noise figure. From Equation (4.18), the output noise is

$$N_o = GN_i + N_e = FGN_i. \tag{4.23}$$

That is, $$N_o|_{\text{dBm}} = \text{NF} + G|_{\text{dB}} + N_i|_{\text{dBm}} \quad . \tag{4.24}$$

So if the input noise of an amplifier with a gain of 20 dB and a noise figure of 3 dB is -90 dBm, then the noise at the output of the amplifier is -67 dBm. This analysis of output noise is only correct if the input source is the equivalent of a resistor held at standard temperature, T_0. So the analysis would not apply if the input to the system was an antenna pointed into the sky (away from cosmic objects such as the sun and moon) which would have a noise temperature of 4 K (due to the cosmic background radiation) plus antenna and cable noise.

EXAMPLE 4.2 Effective Noise Temperature

An antenna with a noise temperature of 50 K is connected to an amplifier with a bandwidth of 20 MHz, a noise figure of 3 dB, and a gain of 10 dB. What is the effective noise temperature at the output of the amplifier?

Solution:

The output effective noise temperature, T_{eff}, is the temperature of a resistor that would produce the same available noise power, N_o, as the output available noise power of the amplifier. The amplifier adds excess noise, N_e, to the input noise, N_i, that is amplified by the amplifier. So to determine N_o, N_e must be determined first. This is obtained from the noise figure, NF, which is defined in terms of noise power of a resistor at the input with a temperature of $T_0 = 290$ K.

The noise factor $F = 10^{\text{NF}} = 1.995$ and from Equation (4.22)

$$\begin{aligned} N_e = (F-1)GkT_0B &= (1.995-1)\cdot 10^{10/10}\cdot(1.3807\cdot^{-23}\ \text{J/K})\cdot(290\ \text{K})\cdot(20\cdot 10^6\ \text{Hz}) \\ &= 7.969\cdot 10^{-13}\ \text{W}. \end{aligned} \tag{4.25}$$

The input noise of a resistor at 50 K is

$$N_i = kTB = (1.3807\cdot^{-23}\ \text{J/K})\cdot(50\ \text{K})\cdot(20\cdot 10^6\ \text{Hz}) = 1.381\cdot 10^{-14}\ \text{W}. \tag{4.26}$$

So the total output noise power is

$$N_o = GN_i + N_e = 10^{10/10}\cdot 1.381\cdot 10^{-14} + 7.969\cdot 10^{-13} = 9.350\cdot 10^{-13}\ \text{W}. \tag{4.27}$$

Thus the effective noise temperature at the output of the amplifier is

$$T_{\text{eff}} = \frac{N_o}{kB} = \frac{9.350\cdot 10^{-13}\ \text{W}}{(1.3807\cdot^{-23}\ \text{J/K})\cdot(20\cdot 10^6\ \text{Hz})} = 3386\ \text{K}. \tag{4.28}$$

EXAMPLE 4.3 Noise Figure of an Attenuator

What is the noise figure of a 20 dB attenuator in a 50 Ω system?

Solution:

Denoting the attenuator as being in a 50 Ω system indicates that an appropriate circuit model to use in the analysis consists of the attenuator driven by a generator with a 50 Ω source impedance, and the attenuator drives a 50 Ω load. Also, the input impedance of the terminated attenuator is 50 Ω, as is the impedance looking into the output of the attenuator when it is connected to the source. The key point is that the noise coming from the source is the noise thermally generated in the 50 Ω source impedance, and this noise is equal to the noise that is delivered to the load, as the impedance presented to the load is also 50 Ω. So the input noise, N_i, is equal to the output noise:

$$N_o = N_i. \tag{4.29}$$

The input signal is attenuated by 20 dB (= 100), so

$$S_o = S_i/100, \tag{4.30}$$

and thus the noise factor is

$$F = \frac{SNR_i}{SNR_o} = \frac{S_i}{N_i}\frac{N_o}{S_o} = \frac{S_i}{N_i}\frac{N_i}{S_i/100} = 100 \tag{4.31}$$

and the noise figure is

$$\text{NF} = 20 \text{ dB}. \tag{4.32}$$

THus the noise figure of an attenuator (or a filter) is just the loss of the component. This is not true for amplifiers of course, as there are other sources of noise, and the output impedance of a transistor is not a thermal resistance.

4.3.1 Noise in a Cascaded System

Section 4.3 developed the noise factor and noise figure measures for a two-port. This result can be generalized for a system. Considering the second stage of the cascade in Figure 4-5, the excess noise at the output of the second stage, due solely to the noise generated internally in the second stage, is

$$N_{2e} = (F_2 - 1)kT_0BG_2. \tag{4.33}$$

Then the total noise power at the output of a two-stage cascade is

$$\begin{aligned} N_{2o} &= (F_2 - 1)kT_0BG_2 + N_{o,1}G_2 \\ &= (F_2 - 1)kT_0BG_2 + F_1kT_0BG_1G_2. \end{aligned} \tag{4.34}$$

This relies on the reasonable assumption that the excess noise added in one stage is uncorrelated to the excess noise from other stages as well as being uncorrelated to the input noise. Thus powers can be added. The second term

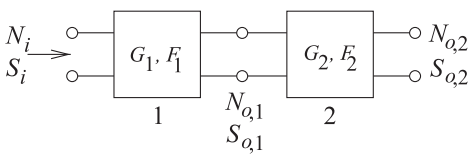


Figure 4-5: Cascaded noisy two-ports.

in Equation (4.34) is the noise output from the first stage amplified by the second stage with gain G_2.

Generalizing the above result yields the total noise power at the output of the mth stage:

$$N_{mo} = \sum_{n=2}^{m} \left[(F_n - 1)\, kT_0 B \prod_{j=2}^{n} G_j \right] + F_1 kT_0 B \prod_{n=1}^{m} G_n. \tag{4.35}$$

Thus an m-stage cascade has a total cascaded system noise factor of $F^{\mathrm{T}} = N_{mo}/\left(G^{\mathrm{T}} N_{1i}\right)$, with G^T being the total cascaded available gain and N_{1i} being the noise power input to the first stage. In terms of the parameters of individual stages, the total system noise factor is

$$F^{\mathrm{T}} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \frac{F_4 - 1}{G_1 G_2 G_3} + \cdots, \tag{4.36}$$

that is,

$$F^{\mathrm{T}} = F_1 + \sum_{n=2}^{m} \frac{F_n - 1}{\prod_{i=2}^{n} G_{i-1}}. \tag{4.37}$$

The similarly named Friis transmission formula refers to antenna systems.

This equation is known as **Friis's formula** [13].

EXAMPLE 4.4 Noise Figure of Cascaded Stages

Consider the cascade of a differential amplifier and a filter shown in Figure 4-6.

(a) What is the midband gain of the filter in decibels? Note that IL is insertion loss.
(b) What is the midband noise figure of the filter?
(c) The amplifier has a gain G_1 = 20 dB and a noise figure of 2 dB. What is the overall gain of the cascade system in the middle of the band? Express your answer in decibels.
(d) What is the noise factor of the cascade system?
(e) What is the noise figure of the cascade system?

Solution:

(a) $G_2 = 1/\mathrm{IL}$, thus $G_2 = -3$ dB.
(b) For a passive element, $\mathrm{NF}_2 = \mathrm{IL} = 3$ dB.
(c) $G_1 = 20$ dB and $G_2 = -3$ dB, so $G_{\mathrm{TOTAL}} = G_1|_{\mathrm{dB}} + G_2|_{\mathrm{dB}} = 17$ dB.
(d) $F_1 = 10^{\mathrm{NF}_1/10} = 10^{2/10} = 1.585$, $F_2 = 10^{\mathrm{NF}_2/10} = 10^{3/10} = 1.995$, $G_1 = 10^{20/10} = 100$, and $G_2 = 10^{-3/10} = 0.5$. Using Friis's formula

$$F_{\mathrm{TOTAL}} = F_1 + \frac{F_2 - 1}{G_1} = 1.585 + \frac{1.995 - 1}{100} = 1.594. \tag{4.38}$$

(e) $\mathrm{NF}_{\mathrm{TOTAL}} = 10 \log_{10}(\mathrm{F}_{\mathrm{TOTAL}}) = 10 \log_{10}(1.594) = 2.03$ dB.

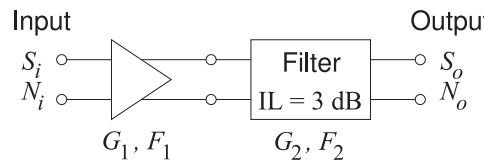


Figure 4-6: Differential amplifier followed by a differential filter.

EXAMPLE 4.5 Noise Figure of a Two-Stage Amplifier

Consider a room-temperature (20°C) two-stage amplifier where the first stage has a gain of 10 dB and the second stage has a gain of 20 dB. The noise figure of the first stage is 3 dB and the second stage is 6 dB. The amplifier has a bandwidth of 10 MHz.

(a) What is the noise power presented to the amplifier in 10 MHz?
(b) What is the total gain of the amplifier?
(c) What is the total noise factor of the amplifier?
(d) What is the total noise figure of the amplifier?
(e) What is the noise power at the output of the amplifier in 10 MHz?

Solution:

(a) Noise power of a resistor at room temperature is -174 dBm/Hz (or more precisely -173.86 dBm/Hz at 293 K). In 10 MHz the input noise power is $N_i = -173.86 \text{ dBm} + 10\log(10^7) = -173.86 + 70 \text{ dBm} = -103.86 \text{ dBm}$.
(b) Total gain $G^{\text{T}} = G_1 G_2 = 10 \text{ dB} + 20 \text{ dB} = 30 \text{ dB} = 1000$.
(c) $F_1 = 10^{\text{NF}_1/10} = 10^{3/10} = 1.995$, $F_2 = 10^{\text{NF}_2/10} = 10^{6/10} = 3.981$. Using Friis's formula, the total noise figure is $F^{\text{T}} = F_1 + \dfrac{F_2 - 1}{G_1} = 1.995 + \dfrac{3.981 - 1}{10} = 2.393$.
(d) The total noise figure is $\text{NF}^{\text{T}} = 10\log_{10}(\text{F}^{\text{T}}) = 10\log_{10}(2.393) = 3.79$ dB.
(e) Output noise power in 10 MHz bandwidth is $N_o = F^{\text{T}} k T_0 B G^{\text{T}} = (2.393) \cdot (1.3807 \cdot 10^{-23} \cdot \text{J} \cdot \text{K}^{-1}) \cdot (293 \text{ K}) \cdot (10^7 \cdot \text{s}^{-1})(1000) = 9.846 \cdot 10^{-11} \text{ W} = -70.07 \text{ dBm}$. Alternatively, $N_o|_{\text{dBm}} = N_i|_{\text{dBm}} + G^{\text{T}}\big|_{\text{dB}} + \text{NF}^{\text{T}} = -103.86 \text{ dBm} + 30 \text{ dB} + 3.79 \text{ dB} = -70.07 \text{ dBm}$.

4.3.2 Measurement of Noise Figure

The Y-factor method is the basis of modern microwave automatic noise figure measurement systems. The technique involves measuring the noise power at the output of a device under test (DUT) when two different noise sources are attached to the input of the DUT [14]. The manual form of the Y-factor method is used at millimeter-wave frequencies. The method is dependent on the accuracy of gain measurement, the ability to generate precise levels of excess noise power, and the sensitivity of noise power measurement. Gain and noise power measurement are subtly different, with gain generally a coherent measurement while the noise power measurement is necessarily incoherent. In both automatic and manual systems, the measurement setup is a cascade system in which the DUT is the first stage and the test set is the last and usually second set. The noise contribution of the test set is only negligible if the gain of the DUT is high.

Under matched conditions, the available power gain of the DUT is

$$G = \frac{S_o}{S_i}, \tag{4.39}$$

so signal power can be eliminated from the expression for the noise factor by combining Equations (4.15) and (4.39) :

$$F = \frac{\text{SNR}_i}{\text{SNR}_o} = \frac{S_i}{N_i}\frac{N_o}{S_o} = \frac{N_o}{N_i G}. \tag{4.40}$$

The output noise is larger than the amplified input noise because of the noise inserted by the DUT. Denoting the component of the output noise power due

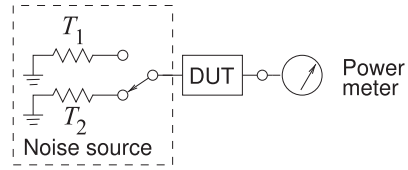


Figure 4-7: Noise figure test setup for determining the noise figure of a DUT.

solely to the DUT by N_D, the output noise power is

$$N_o = N_i G + N_D. \tag{4.41}$$

The final component of the development is noting that the input noise power is related to the temperature of the input match so that $N_i = kTB$ where k is the Boltzmann constant and B is the measurement bandwidth. Conventionally F is referenced to standard temperature $T_0 = 290$K (specifically the input noise temperature is T_0) and so

$$N_D = kT_0BG(F-1). \tag{4.42}$$

In the Y-factor method, two noise sources with known noise temperatures T_1 and T_2 (with $T_2 > T_1$) are applied to the input of the DUT and the corresponding output noise powers N_1 and N_2 measured (see Figure 4-7). In the laboratory a calibrated **noise source**, such as a reverse-biased diode in avalanche, is used to produce the noise source at T_2. The other temperature, T_1, is often obtained by turning the noise source off so that $T_1 = T_0$, the ambient temperature. This leads to the Y factor, which is defined as

$$Y = N_2/N_1. \tag{4.43}$$

In the off state, that is when $T_1 = T_0$, the output noise power in the off state is called the "off power":

$$N_1 = kT_0BG + N_D = kT_0BG + kT_0BG(F-1). \tag{4.44}$$

The second noise source, with noise temperature T_2, produces calibrated excess noise and the noise power under these conditions is called the "on power":

$$N_2 = kT_2BG + N_D = kT_2BG + kT_0BG(F-1). \tag{4.45}$$

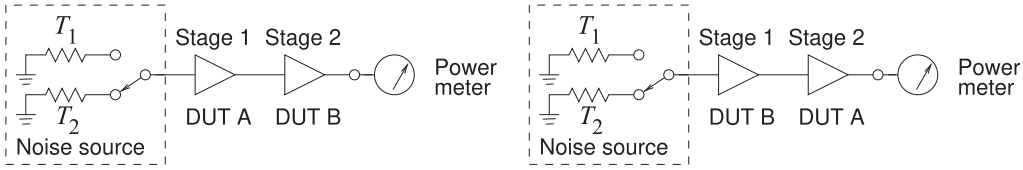
Combining Equations (4.42)–(4.45) yields

$$F = \frac{T_2 - T_0}{T_0(Y-1)}. \tag{4.46}$$

Expressing Equation (4.47) in decibels and integrating (perhaps through measurement) over the system bandwidth yields the noise figure,

$$\mathrm{NF} = 10\log(F) = \mathrm{ENR_{dB}}–10\log(Y-1), \tag{4.47}$$

where $\mathrm{ENR_{dB}} = 10\log[(T_2 - T_0)/T_0]$ is the excess noise ratio in decibels of the calibrated noise source.



(a) Forward cascade (b) Reverse cascade

Figure 4-8: Y-factor test set actively incorporating the second-stage contribution effect.

4.3.3 Measurement of Noise Temperature

If the noise figure of an amplifier replacing the DUT is known, then the noise temperature of a source can be determined. From Equation (4.46),

$$T_2 = FT_0(Y - 1) + T_0. \tag{4.48}$$

In a laboratory application the amplifier could be cryogenically cooled and then there is negligible excess noise contribution by the amplifier so that $F \approx 1$ and $T_2 = YT_0F$.

4.3.4 Measuring the Noise Figure of Low Noise Devices

One of the factors that affects the accuracy of noise figure determination is noise originating in the measurement test set. This leads to an error sometimes referred to as the second-stage contribution effect [15]. This is a particularly important issue when measuring the noise figure of low-gain devices, as then the noise contribution of the second stage can be significant. One approach to minimizing this error is the insertion of a high-gain low-noise amplifier, referred to as the instrumentation amplifier, between the DUT and the test set. The noise figure of the instrumentation amplifier should be known precisely if the error it introduces is to be removed from the raw noise figure measurement.

The Y-factor method relies on the instrumentation amplifier having a much lower noise figure than the DUT. When this cannot be achieved, the best that can be done is to use two (almost) identical amplifiers with the same gain and noise characteristics. The extended Y-factor technique described in this section utilizes two DUTs in a two-stage cascade first with one arrangement of the DUTs and then with the reverse cascade [16]. The technique makes use of the cascaded noise factor operation twice. Figure 4-8 illustrates the test setup with the two possible arrangements of the cascaded DUTs. In Figure 4-8 the power meter is configured to measure noise power (normalized to a 1 Hz bandwidth).

The individual noise factors of the DUTs for a cascaded system with DUT A followed by DUT B, the forward cascade, are denoted by F_{1A} and F_{2B}, and those for the reverse cascade with DUT B followed by DUT A are F_{1B} and F_{2A}. The corresponding gains of the stages are G_{1A}, G_{2B} for the forward cascade, and G_{1B}, G_{2A} for the reverse cascade. Here the first subscript refers to the position in the cascade (either first or second stage), and the second subscript identifies the particular DUT (either A or B). Finally, the total noise factors of the two-cascaded systems are denoted F_A^{T} and F_B^{T} according to

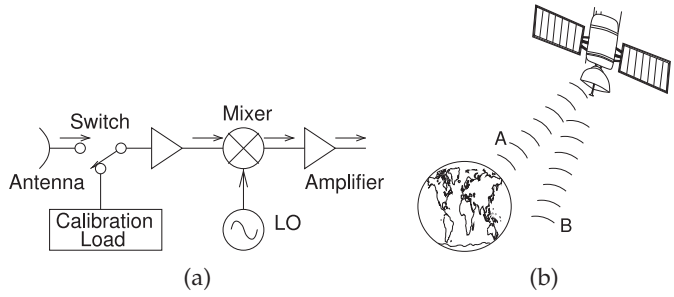


Figure 4-9: Radiometer: (a) heterodyne architecture showing calibration switch; and (b) satellite radiometer that switches between two beams, one oriented at the region being monitored and the other at much colder space that serves as calibration.

whether DUT A or DUT B is the first stage. Using Friis's formula for the forward cascade, Equation (4.37) can be written as,

$$F_A^{\mathrm{T}} = F_{1A} + (F_{2B} - 1)/G_{1A}, \tag{4.49}$$

and for the reverse cascade

$$F_B^{\mathrm{T}} = F_{1B} + (F_{2A} - 1)/G_{1B}. \tag{4.50}$$

The technique presumes that the parameters of the DUTs are invariant of their position in the cascade so that $F_{1A} = F_{2A} = F_A$ and $F_{1B} = F_{2B} = F_B$ as well as $G_{1A} = G_{2A} = G_A$ and $G_{1B} = G_{2B} = G_B$. Equations (4.49) and (4.50) can now be solved simultaneously for the unknown noise factors of the two stages:

$$F_B = \left[F_B^{\mathrm{T}} G_A G_B - G_A(1 - F_A^{\mathrm{T}}) - 1\right]/(G_A G_B) \tag{4.51}$$

and

$$F_A = \left[F_A^{\mathrm{T}} G_A - F_B + 1\right]/(G_A) \tag{4.52}$$

from the measured F_A^{T}, F_B^{T}, G_A, and G_B. So with the gains of the two stages measured independently, the noise factors of the two stages can be determined from the measured noise factors of the forward and reverse cascades.

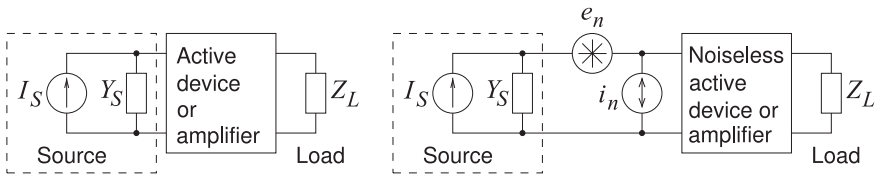
In the special situation of matched DUTs where the noise and gain of the two stages are identical (so that $F_A = F_B = F$ and $G_A = G_B = F$) and $F_A^{\mathrm{T}} = F_B^{\mathrm{T}} = F^{\mathrm{T}}$, then the calculations simplify to yield the noise factor of each stage:

$$F = \frac{GF^{\mathrm{T}}}{G^2 + 1}. \tag{4.53}$$

One of the assumptions of the extended Y-factor method is that the gain and noise of the stages are invariant with respect to the order of the stages in the cascade. Any departure will result in an error. A technique that reduces the sensitivity to placement is to use small attenuators at the inputs and outputs of the stages. The noise contributions can then be removed from the final measured result.

4.3.5 Radiometer System

A radiometer, as shown in Figure 4-9(a), measures the power in EM radiation predominantly at microwave frequencies and most commonly measures noise. Radiometers are used in remote sensing and radio astronomy,



(a) Noisy active device. (b) Noise-free active device.

Figure 4-10: Amplifier model for noise factor calculation: (a) with noisy active device or amplifier where I_S is the source generator and Y_S is the Norton equivalent admittance of the source; and (b) with the noisy active device or amplifier replaced by the noise voltage, e_n, and noise current, i_n, (the noise sources), and a lossless active device or amplifier.

especially by satellites and aircraft. Understanding the physical process that creates uncorrelated radiation at different frequencies enables vegetation, air and sea temperatures, ice coverage, ocean salinity, and other surface and atmospheric sources to be identified from the spectrum captured by a radiometer. Radiometers monitor discrete windows of the spectrum, particularly at frequencies corresponding to molecular resonances. A radiometer includes a mechanism for rapid calibration, such as a **Dicke switch**, quickly switching between the object being observed and another object serving as a calibrated noise source. In aircraft and on land, the calibration sources are resistors, often held at low temperatures. A satellite-based radiometer is shown in Figure 4-9(b), and instead of a Dicke switch, the calibration is obtained by switching the antenna beam between the observed region (Beam A) and an empty area of space (Beam B). With a Dicke switch or antenna beam switching, it is possible to achieve better than 0.01 K of resolution.

4.3.6 Noise Figure of a Two-Port Amplifier

The parameters that define the noise figure of two-port amplifiers were set forth by Haus et al. [12] in 1960 and are used by microwave transistor manufacturers in their datasheets. The amplifier model used in noise factor computation is shown in Figure 4-10.

Figure 4-10 has the active device or amplifier, which is noisy, represented as a two-port with a Norton equivalent source, which will have its own noise, and a load. The noisy active device or amplifier can be replaced by a noiseless two-port with a voltage noise source, e_n, and a current noise source, i_n, at the front of the two-port. Any linear noisy two-port can be represented by a noise free two-port with noise sources. With an active device or amplifier there will be many internal noise sources but it is the noise sources at the input, prior to gain, which are most significant. For an active device or amplifier the voltage and current sources will be at least partially correlated and so the source admittance is important in determining how they combine. A worst case would be when they combine to present the maximum possible noise at the input of the active device or amplifier. It is also possible to adjust the source admittance so that the noise sources combine to present the minimum noise to the active device or amplifier. This effect is captured in the noise

factor of the amplifier in Figure 4-10. The noise factor of this amplifier is [12]

$$F = F_{\min} + \frac{r_n}{g_s}\left|y_s - y_{\text{opt}}\right|^2, \tag{4.54}$$

where $r_n = (R_n/Z_0)$ is called the equivalent noise resistance of the two-port and $F_{\min}$ is the minimum noise factor obtained by adjusting tuners at the input of the amplifier to present all possible values of Y_S to the input of the amplifier. The normalized admittance presented by the tuners at $F_{\min}$ is y_{opt}. With $y_s(= Y_S/Z_0)$ and $g_s = \Re\{y_s\}$ being the actual normalized admittance and conductance, respectively, Equation (4.54) enables the noise factor to be calculated for an actual design. The parameters in Equation (4.54) describe the effect of internal amplifier noise sources and how they are correlated.

More commonly the noise parameters are reported in terms of reflection coefficients rather than admittance. The source reflection coefficient, Γ_s, comes from

$$y_s = \frac{1-\Gamma_s}{1+\Gamma_s} \tag{4.55}$$

and the optimum source reflection coefficient, Γ_{opt}, comes from

$$y_{\text{opt}} = \frac{1-\Gamma_{\text{opt}}}{1+\Gamma_{\text{opt}}}. \tag{4.56}$$

Substituting these into Equation (4.54) results in

$$F = F_{\min} + \frac{4r_n\left|\Gamma_s - \Gamma_{\text{opt}}\right|^2}{\left(1-|\Gamma_s|^2\right)\left|1+\Gamma_{\text{opt}}\right|^2}. \tag{4.57}$$

Together $F_{\min}$, r_n, and Γ_{opt} are called the noise parameters of a device and must be measured. The noise parameters of a pHEMT transistor are given in Table 4-1.

In general, the design for best noise performance does not yield the best gain. The reduction in gain is usually small however. Designing the input and output matching networks of an amplifier to be conjugately matched yields maximum amplifier gain. For best noise performance, however, the input matching network is not conjugately matched and instead the input reflection coefficient looking into the matching network from the active device is Γ_{opt}. As a result, maximum gain is not obtained. The interpretation of why this is necessary is that a particular mismatch minimizes the combined noise contributions of partially correlated internal active device noise sources. In modern RF and microwave design, however, if noise performance is of concern, two or more amplifier stages are used, with the first stage designed for optimum noise performance and subsequent stages designed to obtain the required gain.

4.4 Oscillator Noise

Noise on oscillators has a particular characteristic that affects the use of oscillators in RF systems. When oscillating, the frequency of oscillation varies slightly and this variation is captured as phase noise.

4.4.1 Observations of Noise Spectra of Oscillators

In the frequency domain, noise is characterized by its power spectral density, S, which is the noise power in a specified bandwidth. It is typical to use 1 Hz as the reference bandwidth and then S is expressed in units of watts per hertz or, more commonly, as dBm per hertz. Some types of noise are proportional to the RF signal or carrier level and are expressed relative to the level of the carrier as dBc/Hz ($10\log$ of the ratio of the noise power per hertz relative to the power of the carrier). Noise can be decomposed into a phase component, S_φ, and an amplitude component, S_a. Then the observed phase fluctuations, as classified in [18], are as shown in Figure 4-11. There are also amplitude fluctuations that have a similar characteristic. For example, flicker amplitude noise will also have an f^{-1} dependence. However, the phase component generated by electronic circuits gets the most attention as small amplitude variations are quenched by device nonlinearities. In Figure 4-11, the log of

Frequency (GHz)	$F_{\min}$	$\mathrm{NF}_{\min}$ (dB)	Γ_n	$\angle\Gamma_n$ degrees	$r_n/50$ (Ω)
0.90	1.07	0.29	0.747	15.70	0.165
1.80	1.09	0.38	0.623	24.95	0.176
2.40	1.11	0.44	0.795	37.45	0.158
2.60	1.11	0.47	0.640	47.15	0.159
2.80	1.12	0.49	0.670	47.90	0.160
3.20	1.13	0.53	0.617	51.20	0.156
4.00	1.15	0.61	0.542	68.70	0.141
5.00	1.18	0.72	0.465	85.00	0.120
5.50	1.19	0.77	0.431	91.10	0.114
6.00	1.21	0.83	0.366	101.15	0.107
7.00	1.24	0.93	0.262	122.10	0.096
8.00	1.27	1.04	0.188	153.60	0.100
9.00	1.30	1.14	0.135	−165.60	0.121
10.00	1.33	1.25	0.162	−126.80	0.138
11.00	1.37	1.35	0.183	−85.95	0.187
12.00	1.40	1.46	0.270	−68.40	0.239
13.00	1.43	1.57	0.343	−50.25	0.355
14.00	1.47	1.67	0.431	−43.95	0.461
15.00	1.51	1.78	0.573	−25.80	0.604

Table 4-1: Noise parameters of an enhancement-mode pHEMT transistor model FPD6836P70 [17]. $F_{\min}$ is the minimum noise factor, $\mathrm{NF}_{\min}$ is the minimum noise figure, and $\Gamma_n = \Gamma_{\mathrm{opt}}$ is the source reflection coefficient yielding $F_{\min}$.

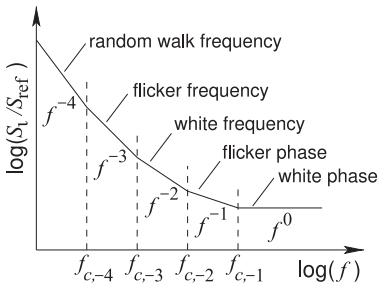


Figure 4-11: Wideband noise showing only phase and frequency noise. S_φ is the spectral density of the phase noise and f is frequency.

the phase noise power spectral density relative to a reference level is plotted as a function of the log of frequency. The graph does not go all the way down to DC because the frequency axis is logarithmic. This does not imply that the noise power goes to infinity[1] as, of course, the bandwidth is becoming smaller and smaller at low $\log(f)$.

Figure 4-11 is the usual way to plot noise, as straight-line regions are clearly observed experimentally and in these regions the noise power spectral density varies with frequency as $1/f^n$, where n is a positive integer. Sometimes one or more of these regions will be missing, and this is interpreted as the crossover frequencies, the $f_{c,-n}$s in Figure 4-11, being out of order.

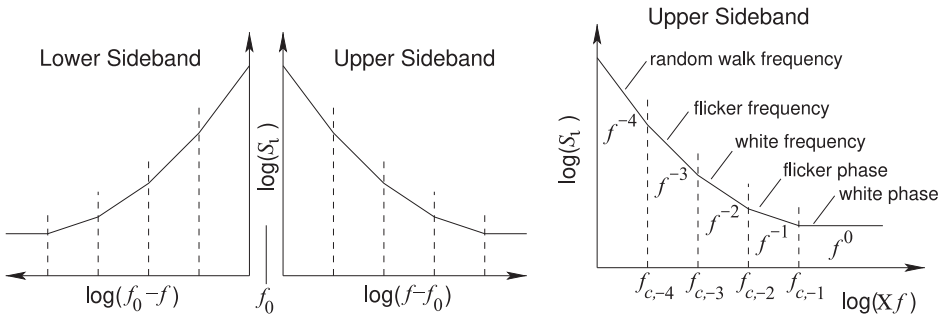
The observed noise characteristic (Figure 4-11) is intriguing. Why are there straight line regions and what are the physical processes that produce noise with such characteristics? The straight-line regions are in contrast to what could, perhaps, be expected to be a smooth continuous transition from a low spectral density region to a high spectral density region.

Names have been given to the regions [18], with the first being white phase noise, noise with no frequency dependence (i.e., a frequency dependence of f^0). This noise has a mean and a variance but no higher (i.e., third and above) moment. In the time domain this noise has (actually assumed to have) a Gaussian distribution. This is the form of noise that can be translated between the time and frequency domains using integer calculus.

The next noise region seen in Figure 4-11 is called the flicker phase noise region with noise spectral density dependent on $1/f$ (or f^{-1}). Traditionally this has been a puzzling noise characteristic and is possibly due to chaotic behavior. The $1/f$ characteristic indicates that there is long-term memory, which is the same as saying that fluctuations (i.e., noise) are correlated to fluctuations at past times. In RF and microwave systems, flicker phase noise is often the most significant noise and sets the noise-related performance limits receiver circuits in particular. Energy consumption is also affected as flicker noise is usually suppressed by increasing oscillator bias currents. Following the $1/f$ region are the white frequency, f^{-2}, flicker frequency, f^{-3}, and random walk frequency regions, f^{-4}. These regions also indicate long-term correlation of fluctuations. It has been shown that at least some of the noise in the f^{-2} region is die to up-converted white noise from near DC [22] and from white noise near harmonics of the oscillating signal [23].

At RF the noise of greatest interest is the noise superimposed on a signal that is amplified, or the noise that is generated by an oscillator. This noise has a white noise component and one or more components with $1/(\Delta f)^n$ dependence, where n is a positive integer and Δf is the offset from the center signal frequency. The noise with $1/(\Delta f)^n$ dependence has been found to

[1] More importantly, the integration, using integer calculus, is not valid. The noise that has the $1/f$-like spectrum is believed to be fractal (i.e., of chaotic origin). Fractals are formally irregular, rough, and nondifferentiable. That is, fractal processes are inaccessible to treatment by integer calculus and thus they do not have a power spectrum [19]. However, it has been shown that such a process when passed through an ideal bandpass filter will become stationary and then have a power spectrum [20, 21]. What this means is that since measurement equipment is band limited, the measured spectrum may appear to approach infinity as the offset goes to zero, but the underlying physical noise process will not. As well, the real bandpass filter of measurement equipment has loss and so a measurement artifact is that noise measurements will level off for smaller and smaller frequency offsets.



(a) Double-sideband noise (b) Single-sideband noise

Figure 4-12: Narrowband noise around the carrier frequency, f_0, of an oscillator showing only phase and frequency noise with $\Delta f = f - f_0$.

be at least partly related to the noise observed at low frequencies. (This is another indication of our lack of a full understanding of noise.)

The phase noise observed on the signal produced by an oscillator is shown in Figure 4-12(a). While the noise appears both above and below the oscillating frequency, so-called double-sideband noise, usually only the upper sideband is plotted, as shown in Figure 4-12(b). The designations of the straight-line regions [18] (see Figure 4-12(b)) correspond to the designations for low-frequency noise (see Figure 4-11). The noise observed on an oscillation signal or on the signal at the output of an amplifier must, necessarily, come from processing of a physical noise source.

The noise on the signal at the output of an oscillator is almost entirely phase fluctuations, as saturation quenches amplitude fluctuations. Phase fluctuations in the frequency domain are characterized by spectral density of the phase fluctuations. If the signal at the output of the oscillator is

$$v(t) = [V_0 + \epsilon(t)] \sin [2\pi f_0 t + \phi(t)] , \tag{4.58}$$

where $\epsilon(t)$ is the amplitude fluctuation and $\phi(t)$ is the phase fluctuation, the spectral density of phase fluctuations is

$$S_\phi(\Delta f) = S_\phi(\Delta\omega) = \text{PSD}\,[\phi(t)] = \frac{E\left[\phi^2(t)\right]}{B}, \tag{4.59}$$

where $\Delta f = \Delta\omega/(2\pi)$ is the frequency offset from the carrier at frequency f_0, PSD refers to power spectral density, $E[\,]$ refers to the estimated value (here the mean of the squared phase), and B is the bandwidth over which the estimate is made (i.e., the bandwidth of the measurement). With bandwidth having the units of hertz, then the units of S_ϕ are $\text{radians}^2/\text{Hz}$. Equation (4.59) includes contributions from both the upper and lower sidebands and so it is a double-sideband measure of phase noise. The preferred measure is the single-sideband phase noise spectral density $\mathcal{L}$ (read as script-L) so that[2]

$$\mathcal{L}(f) = \tfrac{1}{2} S_\phi(f). \tag{4.60}$$

[2] In the past, $\mathcal{L}(f)$ was defined as the single-sideband noise power in a 1 Hz bandwidth divided by the carrier power. This definition has been superseded by Equation (4.60) because of ambiguities in applying the old definition when both amplitude and phase fluctuations are significant.

4.5 Nonlinear Distortion

Distortion imposes a fundamental limit to the efficiency that can be realized in an RF system [24–27]. There must be enough DC power to ensure that signals are processed with no more than the maximum acceptable distortion. Nonlinear distortion originates when the output signal from an amplifier, for example, approaches the extremes of the load line so that the output is not an exact amplified replication of the input signal.

4.5.1 Amplitude and Phase Distortion

For a one-tone signal, the amplitude gain of the signal rolls off as the input power increases, as shown in Figure 4-13(a). This figure plots the power of the output sinewave against the power of the input sinewave. The plot is called the AM-to-AM (**AM-AM**) response (the amplitude-dependent amplitude response) of the amplifier. The AM-AM characteristic is linear at low input powers, but eventually the gain reduces (compresses) and the output power drops below the linear extension of the small-signal response. In Figure 4-13(a), the 1 dB gain compression point is the point where the difference between the extrapolated linear response exceeds the actual gain by 1 dB. $P_{1\text{dB}}$ is the output power at the 1 dB gain compression point and is the single most important metric of distortion, and amplifier designers use $P_{1\text{dB}}$ as a point of reference.

A gain compression of 1 dB corresponds to a signal voltage amplitude reduction of 11% so $P_{1\text{dB}}$ indicates a small but significant reduction in the linear gain of an amplifier. For high orders of modulation this 11% error could mean that the sampled received signal does not match the actual transmitted constellation point. The range of power of a modulated input signal is usually chosen so that the peak power of the envelope is equal to or less than the 1 dB gain compression power. Thus for nearly all modulation formats (a notable exception is FM) the average signal power is backed-off from the 1 dB gain compression power by the **peak-to-mean envelope power ratio** (**PMEPR**) of the modulated signal. With higher-order modulation having a higher PMEPR a greater back-off is required.

At large powers, the parasitic capacitances of the transistors in the amplifier vary the phase of the output signal, and hence phase distortion occurs. Figure 4-13(b) shows what is called the AM-to-PM (**AM-PM**) characteristic which is a measure of the amplitude-dependent phase distortion. For a digitally modulated signal, phase distortion would be

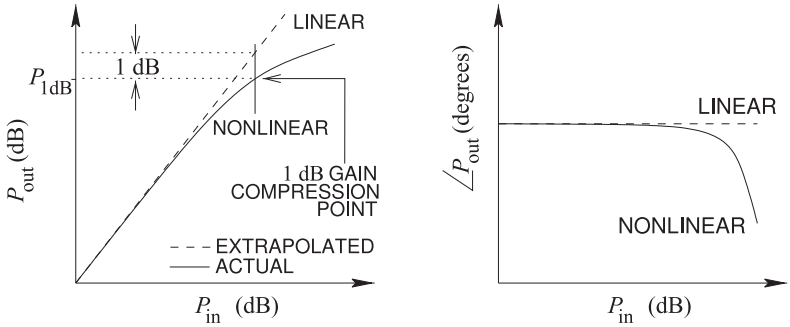


Figure 4-13: Nonlinear effects introduced by RF hardware: (a) amplitude-dependent amplitude (AM-AM) distortion; and (b) amplitude-dependent phase (AM-PM) distortion.

significant if it caused the wrong constellation point to be selected in a receiver. So for 8-PSK modulation a $22.5°$ phase shift would be significant, less if noise is considered. The AM-AM distortion generally occurs before the output phase varies appreciably. This is because the nonlinearity of a transistor is predominantly a current-voltage nonlinearity with phase distortion coming mostly from the nonlinearity of parasitic capacitances.

While Figure 4-13 presents the distortion characteristics for a single sinewave, it has proved to be a reasonable indicator of performance with modulated signals. In particular, with the maximum peak envelope output power being $P_{1\text{dB}}$, in-band distortion is usually acceptable while maximizing high amplifier efficiency [28].

The AM-AM and AM-PM characterizations describe distortion with an amplifier but they are also used with other types of nonlinear subsystems such as mixers.

EXAMPLE 4.6 Amplifier Back-Off

An amplifier has an output power of 10 dBm when the gain of a single tone is compressed by 1 dB. What is the maximum output power of an undistorted QPSK signal with a PMEPR of 3.6 dB?

Solution:

The maximum undistorted output QPSK signal is generally accepted as being when the peak envelope power is equal to the 1 dB gain compression power. In dBm the peak envelope power is greater than the mean signal power by the peak-to-mean envelope power ratio. Thus the QPSK signal power is said to be backed-off from the 1 dB gain compression point by PMEPR. Thus the maximum undistorted output power of the QPSK signal is

$$P_{o,\text{QPSK}} = P_{1\text{dB}}|_{\text{dBm}} - \text{PMEPR}|_{\text{dB}} = 10\ \text{dBm} - 3.6\ \text{dB} = 6.4\ \text{dBm}. \tag{4.61}$$

4.5.2 Gain Compression

The voltage $v_o(t)$ at the output of an amplifier that is biased in the middle of the output current-voltage characteristics (i.e., a Class A amplifier) can be modeled by the first few terms of a Taylor series

$$v_o(t) = a_o + a_1 v_i(t) + a_2 v_i^2(t) + a_3 v_i^3(t) + \ldots, \tag{4.62}$$

where $v_i(t)$ is the input voltage. In a Class A amplifier, the even-order terms are small because distortion at the extremes of the voltage waveform are largely symmetrical, and it is found that the coefficients of the odd-order terms of the Taylor series rapidly decrease so that $|a_1| \gg |a_3| \gg |a_5| \ldots$. Also it is found that a_3 is negative. With a single sinusoidal input voltage at frequency $f_1 = \omega_1/(2\pi)$, $v_i(t) = V_1 \cos(\omega_1 t)$ and the output voltage is

$$\begin{aligned} v_o(t) =& a_0 + a_1 V_i \cos(\omega_1 t) + a_2 V_1^2 \cos^2(\omega_1 t) + a_3 V_i^3 \cos^3(\omega_1 t) + \ldots \\ =& \left(a_0 + \tfrac{1}{2} a_2 V_i^2\right) + \left(a_1 V_i + \tfrac{3}{4} a_3 V_i^3\right) \cos(\omega_1 t) \\ &+ \tfrac{1}{2} a_2 V_i^2 \cos(2\omega_1 t) + \tfrac{1}{4} a_3 V_i^3 \cos(3\omega_1 t) + \ldots . \end{aligned} \tag{4.63}$$

RF and microwave amplifiers are used with bandpass filters and the distortion corresponding to harmonics is easily filtered out. Thus the output, after filtering and removing the DC component is

$$v_o(t) = \left(a_1 V_i + \tfrac{3}{4} a_3 V_i^3\right) \cos(\omega_1 t) = V_o \cos(\omega_1 t), \tag{4.64}$$

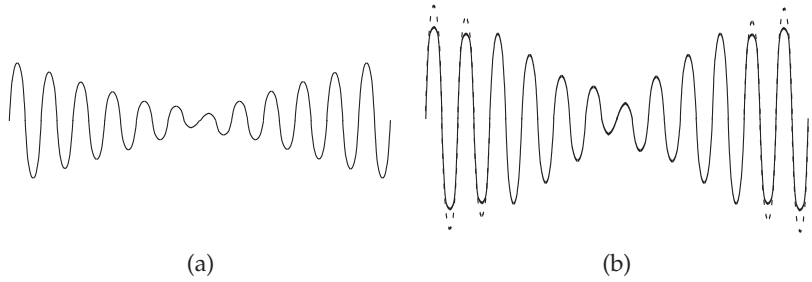


Figure 4-14: A two-tone signal: (a) a two-tone input waveform; and (b) distorted output showing compression (dashed waveform is undistorted).

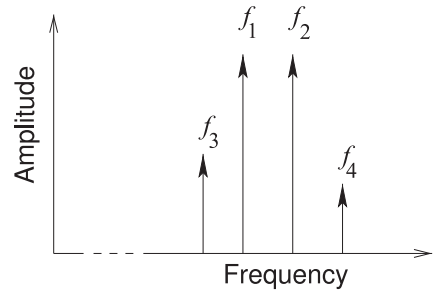


Figure 4-15: Spectrum at the output of a nonlinear amplifier with a two-tone input signal. This is the spectrum of the waveform in Figure 4-14. Numerically $f_3 = 2f_1 - f_2$ and $f_4 = 2f_2 - f_1$.

where $V_o = a_1 V_i + \frac{3}{4} a_3 V_i^3$ is the magnitude of the voltage at the input frequency. So the voltage gain is

$$^{V}G = \frac{V_o}{V_i} = a_1 + \tfrac{3}{4} a_3 V_i^2 \tag{4.65}$$

and the power gain (assuming the input and output impedances are the same) is

$$G = {}^{V}G^2. \tag{4.66}$$

Since a_3 is negative, the gain for very low input voltages, small V_1, is linear with $^{V}G = a_1$ and the gain reduces for larger input signals, resulting in the gain compression, the reduction in the slope of P_{out} versus P_{in}, seen in Figure 4-13(a). For other amplifiers, and particularly for switching amplifiers, many more terms must be considered in the power series expansion, and it is possible that the gain could increase before dropping and eventually saturating. This temporary increase in gain is called gain expansion.

4.5.3 Intermodulation Distortion

A two-tone signal consisting of two sinusoidal signals is a better representation of system performance with modulated signals. A signal that is the linear combination of two sinusoidal signals of equal amplitude is shown in Figure 4-14(a). When this signal is large and is input to an amplifier in what is called a two-tone test, the extremes of the output signal are compressed. This results in the saturated output waveform shown in Figure 4-14(b), where the dashed curve is the undistorted waveform. In the frequency domain this distortion produces additional tones so that this distortion is said to produce **intermodulation products** (**IMPs**), as shown in Figure 4-15. Here the f_1 and f_2 components have the frequencies of the original two-tone input signal. The extra tones in the output, f_3 and f_4, are the intermodulation tones. The tone at $f_3 = 2f_1 - f_2$ is known as the lower **IM3** (or lower third-order intermod) tone and the tone at $f_4 = 2f_2 - f_1$ is known as the upper IM3 tone.

The simplest way to view intermodulation distortion is to consider a two-tone input signal,

$$v_i(t) = V_i \left[\cos(\omega_1 t) + \cos(\omega_2 t)\right], \tag{4.67}$$

where the tones at frequencies $f_1 = \omega_1/(2\pi)$ and $f_2 = \omega_2/(2\pi)$ have equal amplitude V_i. Substituting this signal into the Taylor series expansion in Equation (4.62) leads to the output signal

$$\begin{aligned} v_o(t) &= a_0 + a_1 V_i \cos(\omega_1 t) + a_1 V_i \cos(\omega_2 t) + \tfrac{1}{2} a_2 V_i^2 \left[\cos(\omega_1 t) + \cos(\omega_2 t)\right]^2 \\ &\quad + a_3 V_i^3 \left[\cos(\omega_1 t) + \cos(\omega_2 t)\right]^3 + \ldots \\ &= a_0 + a_1 V_i \cos(\omega_1 t) + a_1 V_i \cos(\omega_2 t) + \tfrac{1}{2} a_2 V_i^2 \left[1 + \cos(2\omega_1 t)\right] \\ &\quad + \tfrac{1}{2} a_2 V_i^2 \left[1 + a_1 \cos(2\omega_2 t)\right] + a_2 V_i^2 \cos(\omega_1 - \omega_2)t + a_2 V_i^2 \cos(\omega_1 + \omega_2)t \\ &\quad + a_3 V_i^3 \left[\tfrac{3}{4} \cos(\omega_1 t) + \tfrac{1}{4} \cos(3\omega_1 t)\right] + a_3 V_i^3 \left(\tfrac{3}{4} \cos(\omega_2 t) + \tfrac{1}{4} \cos(3\omega_2 t)\right] \\ &\quad + a_3 V_i^3 \left[\tfrac{3}{2} \cos(3\omega_1 t) + \tfrac{3}{4} \cos(2\omega_2 - \omega_1)t + \tfrac{3}{4} \cos(2\omega_2 + \omega_1)t\right] \\ &\quad + a_3 V_i^3 \left[\tfrac{3}{2} \cos(3\omega_2 t) + \tfrac{3}{4} \cos(2\omega_1 - \omega_2)t + \tfrac{3}{4} \cos(2\omega_1 + \omega_2)t\right] + \ldots . \end{aligned} \tag{4.68}$$

The component of the output at the first fundamental is

$$V_{o,1} = a_1 V_i + \tfrac{3}{4} a_3 V_i^3 \tag{4.69}$$

and the component of the output at the second fundamental is

$$V_{o,2} = a_1 V_i + \tfrac{3}{4} a_3 V_i^3. \tag{4.70}$$

Note that a_1 is the linear voltage gain of the amplifier:

$$a_1 = V_{o,1}/V_i = V_{o,2}/V_i. \tag{4.71}$$

4.5.4 Third-Order Distortion

RF and microwave amplifiers are typically used with bandpass or lowpass inter-stage matching networks or filters, and so the distortion corresponding to harmonics is filtered out. However, the tones at frequency $f_3 = 2f_1 - f_2 = \omega_3/(2\pi)$ and $f_4 = 2f_1 - f_2 = \omega_4/(2\pi)$ will be within the passband of the amplifier if f_1 and f_2 are close. The appearence of these tones indicates third-order distortion as the distortion is the result of third-order terms in the power series expansion of the two-tone input-output characteristic as given in Equation (4.68). That is, the output after filtering and removing the DC component is

$$v_o(t) = V_{o,1} \cos(\omega_1 t) + V_{o,2} \cos(\omega_2 t) + V_{o,3} \cos(\omega_3 t) + V_{o,4} \cos(\omega_4 t), \tag{4.72}$$

where the component of the output

$$\text{at the first fundamental, } f_1, \text{ is } V_{o,1} = a_1 V_i + \tfrac{3}{4} a_3 V_i^3, \tag{4.73}$$

$$\text{at the second fundamental, } f_2, \text{ is } V_{o,2} = a_1 V_i + \tfrac{3}{4} a_3 V_i^3, \tag{4.74}$$

$$\text{at the lower intermodulation frequency, } f_3, \text{ is } V_{o,3} = \tfrac{3}{4} a_3 V_i^3, \tag{4.75}$$

$$\text{at the upper intermodulation frequency, } f_4, \text{ is } V_{o,4} = \tfrac{3}{4} a_3 V_i^3. \tag{4.76}$$

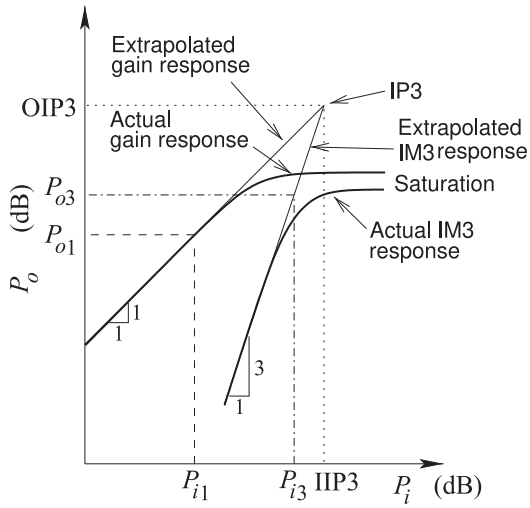


Figure 4-16: Output power versus input power of an amplifier plotted on a logarithmic scale. The IM3 response is a result of two-tone intermodulation, and the input power is the power of each of the two signals that have equal amplitude. Extrapolations of the 1:1 linear fundamental response and the 3:1 third-order intermodulation response intersect at the IP3 point.

Thus the level of the intermodulation tones, the IMD level, increases as the third power of the level of the two-tone input signal. Since the IMD levels vary as the third power of the input tone level (V_i), it is usual to refer to the tones at f_3 and f_4 as third-order intermods, or IM3 tones. The one-tone and IM3 responses are plotted in Figure 4-16 up until saturation, where higher-order terms in the Taylor expansion become important.

The ratio of the amplitude of the intermodulation tones to the amplitude of the input tones (recall that they have equal amplitude) enables the third-order power series coefficient to be calculated. That is

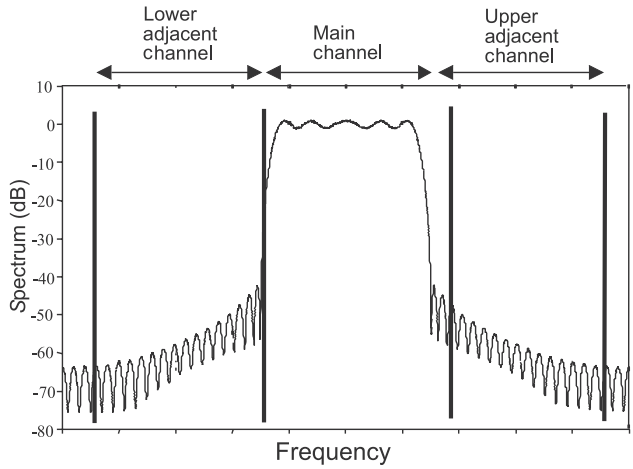
$$a_3 = \frac{4}{3}\frac{V_{o,3}}{V_i^3} = \frac{4}{3}\frac{V_{o,4}}{V_i}. \tag{4.77}$$

The gain and IM3 responses shown in Figure 4-16 are plotted on a logarithmic scale. First consider the gain response, which is plotted for a single sinewave input. At low input power levels the amplifier has linear gain and the output power, P_{o1}, increases in proportion to the input power, P_{i1}, so the gain response has a 1:1 slope. As the input power further increases, the output power saturates primarily because the waveform at the output is constrained by the limits set by the supply and ground rails, but other nonlinearities of the transistor impact the linearity of the response before saturation is reached. The IM3 response (either the level of the f_3 tone or the f_4 tone) in Figure 4-16 is when the levels of two discrete tones are the same. Then typically the responses of the upper and lower IM3 tones are the same unless capacitive effects become important [29–32]. At low levels of the two input tones, each with power P_{i3}, the output power, P_{o3}, at one of the IM3 tones increases as the third power of P_{i3}. Thus on a logarithmic scale the slope of P_{o3} versus P_{i3} is 3:1. As the input power increases the IP3 response eventually saturates. The two simplest characterizations of the nonlinear response of an amplifier are the 1 dB gain compression power, as discussed before, and the intercept of the gain and IP3 responses. This intersection is called the third-order intercept point, or IP3 point (see Figure 4-16). The input power at IP3 is called the input third-order intercept power, or IIP3, and the output is called the output third-order intercept power, or OIP3. If G is the linear power gain, OIP3 $= G \cdot$ IIP3. IIIP3 is mostly used with receivers and OIP3 is used with transmitters.

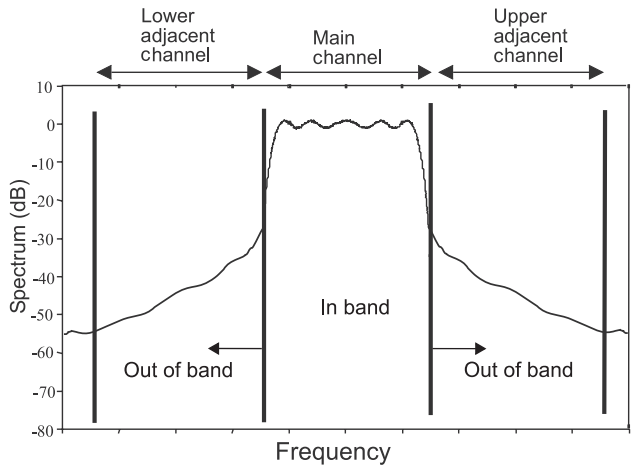
4.5.5 Spectral Regrowth

Distortion with digitally modulated signals consists of in-band and out-of-band distortion. The generation of in-band (within the bandwidth of the digitally modulated signal) intermodulation distortion in an amplifier, or any other nonlinear device such as a mixer, affects the ability to resolve the constellation of a received signal. Thus distortion generated in-band affects the ability to interpret the constellation of the signal and hence causes bit errors. However, the phase of the intermodulation products is of little concern except in cascaded systems where the phase affects how IMD distortion from different stages combines.

Out-of-band distortion is represented in Figure 4-17, where the spectra at the input and output of a nonlinear system are shown. The process that results in increased signal levels in the adjacent sidebands is called spectral regrowth. This distortion is similar to the intermodulation distortion with a two-tone signal. The generation of signals in the adjacent channel affects the function of other radios and the allowable level of these signals is contained in system specifications.



(a) Input spectrum



(b) Output spectrum

Figure 4-17: Spectra at the input and output of an amplifier with of a digitally modulated signal.

4.5.6 Second-Order Distortion

The previous subsection discussed intermodulation distortion and introduced the third-order intercept point, IP3, defined either by the input-referred power of IP3, IIP3, or the output-referred power of IP3, OIP3, to characterize third-order nonlinear performance. The same type of analysis can be used to characterize the second-order nonlinear performance of a microwave module. Second-order distortion leads to the second harmonic of the input tones and to the difference frequency of the input tones. The components of the output response of the amplifier in Equation (4.68) at the second harmonic and difference frequency are:

$$v_o(t) = V_{o,2\text{nd},1}\cos(2\omega_1 t) + V_{o,2\text{nd},2}\cos(2\omega_2 t) + V_{o,\text{diff}}\cos(\omega_2 - \omega_1 t). \quad (4.78)$$

where the amplitude at the output at the

$$\begin{aligned} \text{second harmonic of } f_1 \text{ is } V_{o,2\text{nd},1} &= \tfrac{1}{2}a_2 V_i^2 \\ \text{second harmonic of } f_2 \text{ is } V_{o,2\text{nd},2} &= \tfrac{1}{2}a_2 V_i^2 \\ \text{difference frequency } f_2 - f_1 \text{ is } V_{o,\text{diff}} &= \tfrac{1}{2}a_2 V_i^2. \end{aligned} \quad (4.79)$$

Thus the coefficient of the second-order term in the defining polynomial can be obtained from the amplitude of either of the second harmonics or of the difference tone.

$$a_2 = 2\frac{V_{o,2\text{nd},1}}{V_i^2} = 2\frac{V_{o,2\text{nd},2}}{V_i^2} = 2\frac{V_{o,\text{niff}}}{V_i^2} \quad (4.80)$$

The gain and IM2 responses shown in Figure 4-18 are plotted on a logarithmic scale. At low input power levels the amplifier has a linear gain and initially the gain response has a 1:1 slope. The IM2 response (either the levels of the second harmonic tones or of the difference tone) in Figure 4-18 is when the levels of two discrete tones are the same. At low levels of the two input tones, each with power P_{i2}, the output power, P_{o2}, at one of the harmonic or difference tones increases as the quadratic of P_{i32}. Thus on a logarithmic scale the slope of P_{o2} versus P_{i2} is 2:1. As the input power increases the IP2 response eventually saturates. The simplest characterizations of the second-order nonlinear response of an amplifier is the IP2 response and the IP2 intercept point, the second-order intercept point. The input power at IP2 is called the input second-order intercept power, or IIP2, and the output is called the output second-order intercept power, or OIP2.

4.5.7 Summary

The three simplest characterizations of the performance of an amplifier are the linear gain, the input or output power at the IM2 intercept point, IP2, and the input or output power at the IM3 intercept point, IP3, see Figures 4-18 and 4-19. From these the first three coefficients of a polynomial model of an amplifier can be derived. The first, a_1, comes from the linear gain, see. The second, a_2, comes from the level of the harmonics or difference tones for a two-tone input signal, see Equation (4.80). The third, a_3 comes from the level of third-order intermodulation tones, see Equation (4.77).

4.6 Dynamic Range

While modern communication and radar systems use digitally modulated signals, two-tone signals are used to characterize nonlinearity (usually) and also in manual calculations. At low powers before compression becomes a factor, the fundamental response (i.e., output power versus input power) initially has a 1:1 slope with respect to the input, as shown in Figure 4-19. The third-order intermodulation, IM3, response varies as the cube of the level of input tones when both tones vary by the same amount, as is common in a two-tone test. Thus the IM3 response initially has a 3:1 logarithmic slope with respect to the input. Since the relations are linear in a log-log sense, it is possible to describe the nonlinear performance of an amplifier by a quantity called the **dynamic range** (**DR**) or by the similar **spurious free dynamic range** (**SFDR**). SFDR describes the difference between the level at which a signal is distorted and the level of noise (i.e., the noise floor). DR is similar to SFDR except that the level of the **minimum discernible**

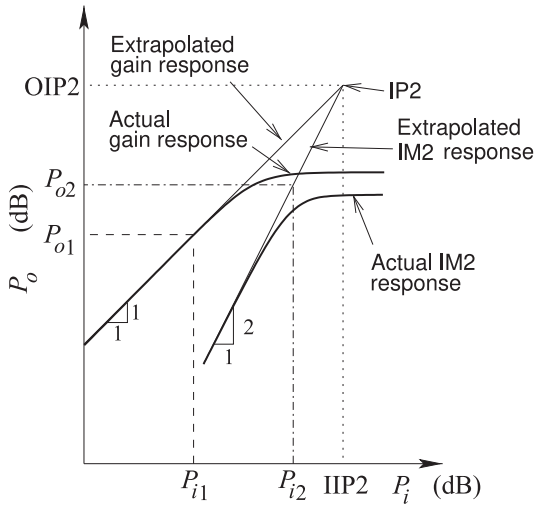


Figure 4-18: Output power versus input power of an amplifier plotted on a logarithmic scale. The IM2 response is a result of two-tone intermodulation (or of harmonic generation), and the input power is the power of each of the two signals that have equal amplitude. Extrapolations of the 1:1 linear fundamental response and the 2:1 second-order response intersect at the IP2 point.

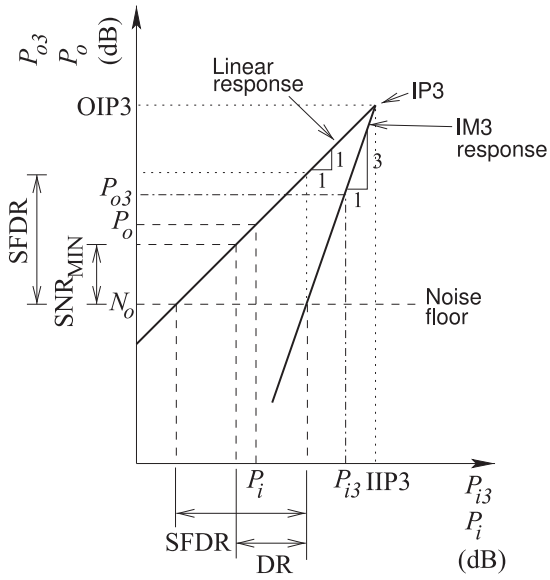


Figure 4-19: Output power versus input power of a stage or system plotted as output power in decibels versus input power in decibels. The IM3 response is a result of two-tone intermodulation, and the input power is the combined power of the two signals that have equal amplitude. Extrapolations of the 1:1 linear fundamental response and the 3:1 third-order intermodulation response intersect at the IP3 point.

signal (**MDS**) (also called the **minimum detectable signal**) is used instead of the noise floor. MDS is higher than the level of the noise floor by the minimum acceptable SNR (SNR_{MIN}). SNR_{MIN} is dependent on the type of modulation, on hardware inadequacies (captured by the implementation margin), on processing gain, and on the error correction coding used in a particular communications protocol.

It is interesting to note that the 1 dB gain compression level has no effect on the dynamic range of microwave circuits and systems. This is because dynamic range is concerned with the ability to detect a signal when it is possible for the desired signal to be masked by spurious signals. Gain compression on its own does not introduce spurious signals so it does not interfere with the ability to detect a small signal.

The dynamic range does not relate to how accurately the sampled received signal matches the constellation diagram of the transmitted digitally modulated signal. Distortion of the constellation diagram is determined by noise, gain compression, and intermodulation distortion.

In the following, an expression for SFDR is developed in terms of input-referenced quantities, and this form of the SFDR is called the input referred SFDR (SFDR_i). A similarly referenced dynamic range (DR_i) is also developed.

Figure 4-19 graphically defines the dynamic ranges. The point of intersection of the extrapolated linear and IM3 responses is called the third-order intercept point (**IP3 intercept**). The point is identified by the output-referred intercept power (OIP3) or by the input-referred IP3 intercept power (IIP3) and these are key parameters in describing the linearity of nonlinear subsystems.

In the linear gain region, the output power P_o versus the input power P_i has a slope of 1:1, so that

$$P_{\text{dBm},i} = P_{\text{dBm},o} - G_{\text{dB}}, \tag{4.81}$$

where G_{dB} is the power gain in decibels. P_o is used here as the output power, with $P_{\text{dBm},o}$ indicating the output power in dBm. P_i and $P_{\text{dBm},i}$ are similarly defined. In terms of input quantities

$$\text{IIP3} = \frac{\text{OIP3}}{G} = \text{IIP3}_{\text{dBm}} = \text{OIP3}_{\text{dBm}} - G_{\text{dB}}, \tag{4.82}$$

where again the dBm subscript indicates that the quantity is expressed in decibels referred to 1 mW.

The nonlinearity of RF active components results in harmonics and intermodulation components. With the narrowband amplifiers of communication and radar systems, output filters conveniently filter out harmonics. However, intermodulation distortion cannot be filtered out, as these components are within the main passband. The intermodulation components are therefore spurious tones. Generally just one of these defines the maximum spurious tone and nearly always it is one of the third-order intermodulation tones resulting from a two-tone input. Consideration of the maximum spurious tone and the noise floor defines the SFDR.

Examining Figure 4-19 leads to the following inequality describing the

linear gain of the amplifier:

$$\frac{\mathrm{OIP3_{dBm}} - P_{\mathrm{dBm},o}}{\mathrm{IIP3_{dBm}} - P_{\mathrm{dBm},i}} = \frac{\mathrm{OIP3_{dBm}} - P_{\mathrm{dBm},o}}{(\mathrm{OIP3_{dBm}} - G_{\mathrm{dB}}) - P_{\mathrm{dBm},i}} = 1. \tag{4.83}$$

The IM3 response is characterized by first introducing an equivalent input power, $P_{\mathrm{dBm,i3}}$ (P_{i3} expressed in dBm), defined as the average power of the two-tone signal that generates an IM3 of power $P_{\mathrm{dBm,o3}}$. Noting that $P_{\mathrm{dBm,o3}}$ varies with a 3:1 logarithmic slope with respect to $P_{\mathrm{dBm,i3}}$, then

$$\frac{\mathrm{OIP3_{dBm}} - P_{\mathrm{dBm,o3}}}{(\mathrm{OIP3_{dBm}} - G_{\mathrm{dB}}) - P_{\mathrm{dBm,i3}}} = 3 \tag{4.84}$$

or

$$P_{\mathrm{dBm,i3}} = \tfrac{1}{3}\left(2 \cdot \mathrm{OIP3_{dBm}} + P_{\mathrm{dBm,o3}} - 3G_{\mathrm{dB}}\right). \tag{4.85}$$

The SFDR can now be defined when the third-order intermodulation product of two-tone excitation is the dominant spurious tone. The SFDR is defined as the difference between P_{i3} and P_i when they produce IM3 and linear output, respectively, that are both equal to the output noise power, N_o (see Figure 4-19); that is, when $P_o = P_{\mathrm{o3}} = N_o$. Replacing $P_{\mathrm{dBm},o}$ in Equation (4.85) with N_o gives

$$P_{\mathrm{dBm,i3}} = \tfrac{1}{3}\left(2 \cdot \mathrm{OIP3_{dBm}} + N_{\mathrm{dBm},o} - 3G_{\mathrm{dB}}\right) \tag{4.86}$$

and

$$P_{\mathrm{dBm},i} = N_{\mathrm{dBm},o} - G_{\mathrm{dB}}. \tag{4.87}$$

Note that the difference between the linear output and the third-order intermodulation reduces as the input power increases above P_{i3}. Thus the output-referred SFDR is

$$\begin{aligned} \mathrm{SFDR_{dB},}_o &= P_{\mathrm{dBm,i3}} - P_{\mathrm{dBm},i} \\ &= \tfrac{2}{3}\mathrm{OIP3_{dBm}} + \tfrac{1}{3}N_{\mathrm{dBm},o} - G_{\mathrm{dB}} - N_{\mathrm{dBm},o} + G_{\mathrm{dB}} \\ &= \tfrac{2}{3}\left(\mathrm{OIP3_{dBm}} - N_{\mathrm{dBm},o}\right). \end{aligned} \tag{4.88}$$

So SFDR is 2/3rds of the range in decibels from the noise intercept to the third order intercept point. A similar development defines the identical input-referred SFDR:

$$\mathrm{SFDR_{dB},}_i = \tfrac{2}{3}\left(\mathrm{IIP3_{dBm}} - N_{\mathrm{dBm},i}\right). \tag{4.89}$$

Note that N_i is the input-referred noise and includes noise applied to the module as well as the noise produced internally in the module and referred to the input. The SFDR provides a combined measure of distortion and noise. However, for usable dynamic range the minimum acceptable SNR must be considered. The minimum SNR ($\mathrm{SNR_{MIN}}$) required is determined by the communication or radar modulation format, error coding, and acceptable BER. So in defining DR, the input power of the desired signal must increase sufficiently to produce an SNR of at least $\mathrm{SNR_{MIN}}$. Since the desired spurious level is still at the noise floor, this implies a direct subtraction in decibels of the desired SNR. Therefore the input-referred third-order dynamic range, preferred for receivers, is

$$\mathrm{DR}_i = \tfrac{2}{3}\left(\mathrm{IIP3_{dBm}} - N_{\mathrm{dBm},i} - \mathrm{SNR_{dB,MIN}}\right) \tag{4.90}$$

and the output-referred dynamic range, preferred for transmitters, is

$$\mathrm{DR}_o = \tfrac{2}{3}\left(\mathrm{OIP3}_{\mathrm{dBm}} - N_{\mathrm{dBm},o} - \mathrm{SNR}_{\mathrm{MIN}}\right). \tag{4.91}$$

The minimum discernible signal at the output is

$$\mathrm{MDS}_{\mathrm{dBm},o} = N_{\mathrm{dBm},o} + \mathrm{SNR}_{\mathrm{dB,MIN}}, \tag{4.92}$$

so the output-referred dynamic range can be written as

$$\mathrm{DR}_o = \tfrac{2}{3}\left(\mathrm{OIP3}_{\mathrm{dBm}} - \mathrm{MDS}_{\mathrm{dBm},o}\right). \tag{4.93}$$

So DR is 2/3rds of the range in decibels from the level of the minimum detectable signal to the third order intercept point. The minimum discernible signal at the input is

$$\mathrm{MDS}_{\mathrm{dBm},i} = N_{\mathrm{dBm},i} + \mathrm{SNR}_{\mathrm{dB,MIN}} \tag{4.94}$$

and the input-referred dynamic range (equal to DR_o) can be written as

$$\mathrm{DR}_i = \tfrac{2}{3}\left(\mathrm{IIP3}_{\mathrm{dBm}} - \mathrm{MDS}_{\mathrm{dBm},i}\right). \tag{4.95}$$

4.7 Passive Intermodulation Distortion

Passive intermodulation distortion (**PIM**) is the term used to describe nonlinear distortion that occurs when passive components are supposedly linear. Just as with nonlinear components such as amplifiers, the generation of spurious tones can swamp small received signals. The typical test of PIM is a two-tone test in which two large equal-amplitude sinusoids are applied to the device under test (DUT) and the level of the intermodulation signals measured. Except for the DUT generating the distortion being passive, distortion appears as described in Section 4.5. PIM is observed in a wide number of situations including coaxial cables [33, 34], microstrip transmission lines [35], attenuators [36], terminations [37, 38], rectangular waveguides [39, 40], antennas, and filters [41, 42].

4.7.1 Sources of PIM

Several causes of PIM have been identified or suggested. These include self-heating and nonlinear junction effects when currents flow from one metal into a dissimilar metal. The concept being that the dissimilar metals form a weak diode-like current-voltage characteristic [39].

The most convenient method for measuring small levels of PIM uses a filter to separate a small distorted intermodulation signal from the two driving tones. The finite filter response requires that the RF tones be at least 1 MHz apart. However, distortion that is less than 1 MHz from a signal is of concern in RF systems. Several measurements have shown that the level of distortion increases rapidly as the distortion component gets closer in frequency to the driving signal [37].

Figure 4-20 shows the third-order PIM of terminations measured in a two-tone test [43]. The PIM level is plotted against the frequency separation of the two tones. Similar results are seen with other components including transmission lines and antennas [35–38]. The low-PIM termination, resulting in Curve (a) in Figure 4-20, is the lowest PIM termination available

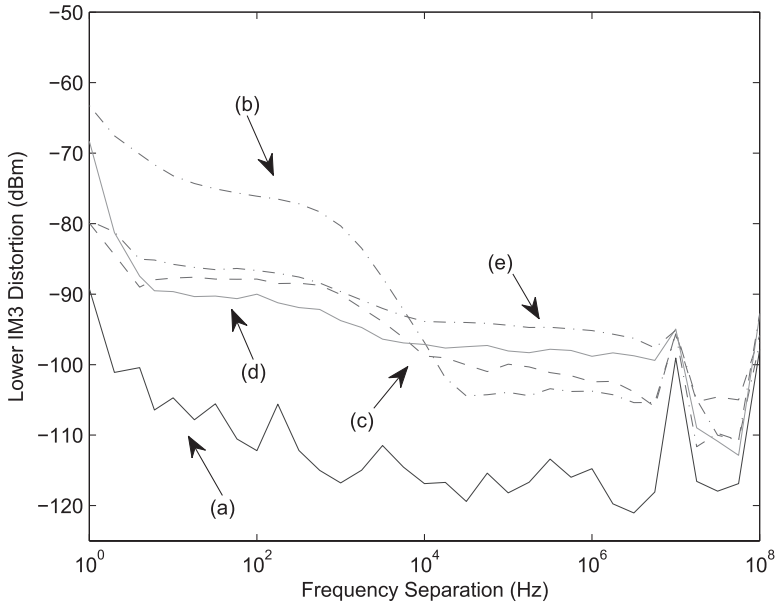


Figure 4-20: Measured third-order passive intermodulation distortion (lower IM3) of common laboratory high-power, finned N-type terminations using a two-tone test at 460 MHz: (a) low-PIM cable-based termination; (b) part PE6097, component A; (c) part PE6097, component B; (d) part PE6035, component A; and (e) part PE6035, component A. Measurement taken with 26 dBm input power for each tone. After [43].

comprising, here, a terminated 100 m long slightly lossy cable with a loss of 0.1 dB/m. The distributed low-level loss spreads the heat generated out over a considerable distance and so there is a negligible electrothermal PIM effect. Curves (b) and (c) plot the measured PIM for two otherwise identical finned N-type connectors. Curves (d) and (e) plot the measured PIM for another pair of connectors. The reason for the discrepancy is unknown. The rapid rise in PIM below 200 Hz tone separation is due to electrothermal effects.

The rapid increase in the level of PIM as the tone separation reduces is due to electrothermal effects. When close in frequency, the two tones produce a beating waveform that periodically heats a conductor, changing its resistance through the thermal coefficient of resistance of the conductor. This change in resistance results in a periodic variation of the current-voltage relation (i.e., Ohm's law) which results in the generation of intermodulation tones. It has been shown that if dissimilar metals and magnetic materials are avoided, the PIM generated is entirely due to electrothermal effects [37].

With tone spacings exceeding a few hundred kilohertz, or in some cases a few megahertz, the PIM is being generated through another mechanism. Several sources of this PIM have been postulated, and while all could be physical sources of PIM, which is most important is not clear. PIM has been shown to be almost entirely due to current density effects so reducing current density is an effective means to reduce PIM [44]. Plausible mechanisms generating such PIM include metal-insulator-metal and metal-metal contact nonlinearities [45], particularly due to surface topography at contacts. For example, it is known that high force applied across contacts reduces PIM. In cellular base stations where the power of signals is very high, it is known to be important to have very tight connections of cables carrying high-power signals.

Tunneling at a metal-insulator-metal contact could produce PIM. This is supposed to be a particular problem with aluminum-oxide-aluminum contacts, as the oxide thickness may be just right, 2 nm, for tunneling to occur [40, 45]. However, this is expected to be a minor contributor to PIM.

Another possible source of PIM is thermionic emission [46]. Thermionic emission is secondary to tunneling and will result in a small increase in tunneling current due to other sources [46].

The metal-to-metal contact of dissimilar metals acts as a weak diode due to the difference in the work functions of the metals. Alignment of Fermi levels requires charge to transfer from the high work function metal to the low work function metal. Since charge transfer has occurred between the metals, a field therefore exists at the interface and so there is a contact potential. It is known that copper, silver, and gold contacts form low PIM contacts and these metals have very similar work functions, supporting the view above. Also at metal-to-metal contacts there is a constriction of current flow due to rough surface topology that exaggerates PIM due to increased current density.

Ferromagnetic materials such as iron, steel, cobalt, and nickel produce high significant PIM [47]. This is also true for ferroelectric and piezoelectric materials.

4.7.2 Summary

There are many possible sources of PIM, and provided that care is used in avoiding dissimilar metals and avoiding ferromagnetic materials, the only confirmed source of RF PIM is the electrothermal self-heating effect [35, 37]. However, it is not always possible to build such components as desirable mechanical, electrical, and packaging requirements necessitate the use of dissimilar elements. Ferromagnetic metals such as nickel, a desirable processing material, produce significant PIM. With dissimilar nonmagnetic materials there is clearly another (unknown) source of PIM at high tone spacings. Even with carefully designed components, PIM exists at large tone spacings, and the main source of PIM in these circumstances has not yet been identified. There is a reasonable confidence that such PIM is a current nonlinearity and not a voltage nonlinearity [44]. So a strategy for reducing PIM is to avoid high current densities, and for electrothermal PIM, to provide rapid dissipation of heat as close to the source of heat as possible. Other sources of PIM-like responses, that is, RF distortion where it is not expected, have been found due to vibration [48, 49] and the transient response of a filter [41, 42].

4.8 Breakdown

Breakdown occurs when the electric field is strong enough that a stray electron strips electrons from atoms or molecules resulting a cascading effect producing a conducting plasma of electrons. There are two types of effects that cause breakdown. One of these is the multipactor effect which occurs at the interface of a metal or dielectric with a vacuum or low pressure gas where electrons are stripped from the metal or dielectric. The other is the corona effect which occurs in a gas, even at fairly low pressures, in which electrons are stripped from atoms or molecules in the gas. The plasma of electrons shorts conducting surfaces at different potential, and in the worst case causes destruction of components. But even before a sufficiently dense plasma is established, effects include noise generation, increase in ohmic losses, and nonlinear signal distortion since breakdown is strongly nonlinear effect the number of free electrons created is a strong nonlinear function of electric field

and thus voltage.

4.8.1 Multipactor Effect

The multipactor effect is due to the secondary emission of electrons when a fast moving electron in a vacuum or low pressure gas impacts a metal or a dielectric. If the energy of the free electron is above a first critical field E_{c1}, an electron in the material will be excited and sometimes will be emitted from the material. The initial electron will penetrate the material by several lattice constants but follow a zig-zag path. If the energy of this electron is below a second critical value E_{c2}, then it will only penetrate one or a few lattice constants and will itself be emitted back into the vacuum (or low pressure gas). If the electron has an initial energy above E_{c2} then it will penetrate the material a considerable distance and will become trapped. The actual material and the surface finish have a significant effect on electron emission.

The number of electrons that leave the surface for each impacting electron is called the **secondary electron emission** yield (**SEY**) and this includes the initial electron which may or may not be emitted. If SEY is greater than one then the number of free electron will avalanche and could eventually lead to device destruction. Before that occurs, there is a balancing effect which limits growth in free electron numbers. Principle among these is the field that is produced by the free electrons themselves which counters the field created by an externally applied voltage potential. The reference geometry used in measuring the multipactor effect is a parallel plate structure with a voltage V between the plates that are separated by a distance d. This geometry generally has the lowest thresholds for the multipactor effect. For other geometries, i.e. non-parallel plate geometries, the high field region where the secondary electrons are created may not be where the plasma, i.e. the region where the free electrons accumulate, is located. Thus the growth in the number of free electrons is curtailed. A further limiting effect is the frequency of the signal applied to the plates where a high frequency signal could reverse limiting the build-up of free electrons. Thus the applied voltage, $V_{\mathbf{multipactor}}$, required to initiate the multipactor effect is a function of d, material, material finish, and frequency f. If there is a free electron between the plates then the electron will accelerate and acquire velocity and thus energy. The voltage required to initiate impact ionization. It is found experimentally that there is a minimum voltage $V_{\mathbf{min}}$ required for the effect to be initiated. Then there is a linear relationship between $V_{\mathbf{multipactor}}$ and the product $f \cdot d$. Materials and material finish also have an impact on ionization For example cleaned mirror-finish silver has a first critical energy, $E_{c1} = 130$ eV whereas silver exposed to the air has $E_{c1} = 20$ eV. Guidelines for the thresholds for multipactor ionization for various materials are given in [50].

4.8.2 Corona Effect

Multipaction describes ionization of materials in a vacuum or in when the environment has a low pressure gas. As the pressure of the gas increases the molecules in the gas will breakdown and this is called the corona effect. If e is an electron and M is a molecule then the corona ionization process is

described by

$$\mathrm{e} + \mathrm{M} \rightarrow \mathrm{e} + \mathrm{e} + \mathrm{M}^{+} \tag{4.96}$$

and the molecule takes on a positive charge after an electron has been stripped off. The competing process which limits the number of free electrons available is when a free electron combines with a molecule and in air these interactions principally involve nitrogen N and oxygen O:

$$\mathrm{e} + \mathrm{N}_2^{+} \rightarrow \mathrm{N} + \mathrm{N} \quad \text{and} \quad \mathrm{e} + \mathrm{O}_2 \rightarrow \mathrm{O}^{-} + \mathrm{O} \tag{4.97}$$

The minimum energy required for the corona effect is heavily dependent on gas pressure. At very low pressures ionization is solely due to multipaction. Above a critical but low pressure multipaction is replaced by the corona effect as as the density of molecules increases it becomes easier to ionize more molecules. The minimum energy required for the corona effect occurs when the pressure in multiples of standard atmospheric pressure is approximately equal to the frequency in gigahertz. Further increase in the pressure reduces the free main path of an electron and fewer free electrons are able to acquire the requisition energy. [50].

4.8.3 Summary

Breakdown can lead to device destruction but even at very low levels it can cause intermodulation products, i.e. passive intermodulation distortion, that obscure small received signals. Breakdown is a particular problem when there are high fields, especially in filters with high energy densities because of resonance. Breakdown can be a problem in basestations, radars, and satellites as these often operate with EM powers of kilowatts to megawatts, and even higher powers when operated with low duty cycles.

4.9 Summary

The use of modules has become increasingly important in microwave engineering. A wide variety of passive and active modules are available and high-performance systems can be realized enabling many microwave systems of low-to-medium volumes to be realized cost effectively and with stellar performance. Module vendors are encouraged by the market to develop competitive modules that can be used in a wide variety of applications. The challenge is to develop a module with high performance and adaptability. Many modules have matched $50\ \Omega$ input and output impedances so that modules can be freely interconnected. Many integrated systems are first prototyped using vendor-supplied modules that are, perhaps, gradually replaced by higher-performance monolithically integrated implementations. Doing so is only justified when unit volumes are very high. The availability of evaluation boards with coaxial connectors for many modules makes it easy to do early design trade-offs. The great majority of RF and microwave engineers either develop modules or use modules to realize RF systems.

The most important part of a system design is achieving the desired functionality largely arrived at by choosing a suitable topology, usually cascaded subsystems, and appropriate functional units (i.e., modules). Then the major concern in design is managing noise, distortion, DC power consumption, cost, and size. The noise and distortion concerns

can be combined together as dynamic range and then the objective is maximizing dynamic range while controlling DC power consumption and using affordable modules. Systems must also be concerned with distortion that originates in supposedly linear elements. Something as simple as a termination can produce PIM through heating and thus resistance changes.

4.10 References

[1] H. Nyquist, "Thermal agitation of electric charge in conductors," *Physical review*, vol. 32, no. 1, p. 110, 1928.

[2] D. Chandler, *Introduction to modern statistical mechanics*. Oxford University Press, 1987.

[3] M. Steer, *Microwave and RF Design, Transmission Lines*, 3rd ed. North Carolina State University, 2019.

[4] H. Nyquist, "Thermal agitation of electric charge in conductors," *Phys. Rev.*, vol. 32, p. 110, 1928.

[5] D. Middleton, *An Introduction to Statistical Communication Theory*. IEEE Press, 1996.

[6] M. Gupta, "Thermal noise in nonlinear resistive devices and its circuit representation," *Proc. of the IEEE*, vol. 70, no. 8, pp. 788–804, Aug. 1982.

[7] A. Hati, D. Howe, F. Walls, and D. Walker, "Merits of pm noise measurement over noise figure: a study at microwave frequencies," *IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 53, no. 10, pp. 1889–1894, Oct. 2006.

[8] H. Hartnagel, R. Katilius, and A. Matulionis, *Microwave Noise in Semiconductor Devices*. Wiley, 2001.

[9] R. Sarpeshkar, T. Delbruck, and C. Mead, "White noise in mos transistors and resistors," *IEEE Circuits and Devices Magazine*, vol. 9, no. 6, pp. 23–29, Nov. 1993.

[10] A. Papoulis and S. Pillai, *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 1994.

[11] Y. M. Blanter and M. Büttiker, "Shot noise in mesoscopic conductors," *Physics Reports*, vol. 336, no. 1/2, pp. 1–166, Sep. 2000.

[12] H. Haus, W. Atkinson, G. Branch, W. Davenport, W. Fonger, W. Harris, S. Harrison, W. McLeod, E. Stodola, and T. Talpey, "Representation of noise in linear twoports," *Proc. of the IRE*, vol. 48, no. 1, pp. 69–74, Jan. 1960.

[13] H. Friis, "Noise figures of radio receivers," *Proc. of the IRE*, vol. 32, no. 7, pp. 419–422, Jul. 1944.

[14] W. Mumford and E. Scheibe, *Noise Performance Factors in Communication Systems*. Horizon House, 1968.

[15] "Noise figure measurement accuracy-the y-factor method," Agilent Technologies, application Note 57-2.

[16] A. Victor and M. Steer, "Improved y factor noise measurement using the second stage contribution to advantage," in *65th ARFTG Conf. Digest, Spring 2005*, Jun. 2005, p. 3.

[17] QORVO, "FPD6836P70 data sheet, low noise high frequency packaged enhancement mode phemt transistor," http://www.qorvo.com.

[18] "IEEE Standard 1139-2008, Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology—Random Instabilities," IEEE, Feb. 2009.

[19] B. Mandelbrot and J. Van Ness, "Fractional brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, no. 4, pp. 422–437, 1968.

[20] G. Wornell, "Wavelet-based representations for the 1/f family of fractal processes," *Proc. of the IEEE*, vol. 81, no. 10, pp. 1428–1450, Oct. 1993.

[21] G. Wornell and A. Oppenheim, *Signal Processing with Fractals: A Wavelet-Based Approach*. Prentice Hall, 1996.

[22] D. Leeson, "A simple model of feedback oscillator noise spectrum," *Proc. of the IEEE*, vol. 54, no. 2, pp. 329–330, Feb. 1966.

[23] A. Hajimiri and T. Lee, "A general theory of phase noise in electrical oscillators," *IEEE J. of Solid-State Circuits*, vol. 33, no. 2, pp. 179–194, Feb. 1998.

[24] W. Pastori, "Effects of DUT mismatch on the noise figure characterization: a comparative analysis of two Y-factor techniques," *Microwave Journal*, vol. 26, no. 4, pp. 50–60, Apr. 1983.

[25] J.-M. Collantes, R. D. Pollard, and M. Sayed, "Effects of dut mismatch on the noise figure characterization: a comparative analysis of two Y-factor techniques," *IEEE Trans. on Instrumentation and Measurement*, vol. 51, no. 6, pp. 1150–1156, Dec. 2002.

[26] U. Rohde and J. Whitake, *Communications Receivers*. McGraw Hill, 2001.

[27] J. Pedro and C. N.B., *Intermodulation Distortion in Microwave and Wireless Circuits*. Norwood, MA, USA: Artech House, Inc., 2003.

[28] J. Sevic and M. Steer, "On the significance of envelope peak-to-average ratio

for estimating the spectral regrowth of an RF/microwave power amplifier," *IEEE Trans. on Microwave Theory and Techniques*, vol. 48, no. 6, pp. 1068–1071, 2000.

[29] N. de Carvalho and J. Pedro, "A comprehensive explanation of distortion sideband asymmetries," *IEEE Trans. on Microwave Theory and Techniques*, vol. 50, no. 9, pp. 2090–2101, Sep. 2002.

[30] A. Walker, M. Steer, and K. Gard, "Capturing asymmetry in distortion of an RF system using a multislice behavioral model," *IEEE Microwave and Wireless Components Letters*, vol. 16, no. 4, pp. 212–214, 2006.

[31] W. Jang, A. Walker, K. Gard, and M. Steer, "Capturing asymmetrical spectral regrowth in RF systems using a multislice behavioral model and enhanced envelop transient analysis," *Int. Journal of RF and Microwave Computer-Aided Engineering*, vol. 16, no. 4, pp. 400–407, 2006.

[32] W. Jang, N. Kriplani, and M. Steer, "Behavioural modelling of amplifier asymmetry in the time domain," *Int. Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 2012.

[33] M. Amin and F. Benson, "Coaxial cables as sources of intermodulation interference at microwave frequencies," *IEEE Trans. on Electromagnetic Compatibility*, vol. EMC-20, no. 3, pp. 376–384, Aug. 1978.

[34] M. Amin and I. Benson, "Nonlinear effects in coaxial cables at microwave frequencies," *Electronics Letters*, vol. 13, no. 25, pp. 768–770, 8 1977.

[35] J. Wilkerson, P. Lam, K. Gard, and M. Steer, "Distributed passive intermodulation distortion on transmission lines," *IEEE Trans. on Microwave Theory and Techniques*, vol. 59, no. 5, pp. 1190–1205, May 2011.

[36] J. Wilkerson, K. Gard, and M. Steer, "Electrothermal passive intermodulation distortion in microwave attenuators," in *36th European Microwave Conf.*, Sep. 2006, pp. 157–160.

[37] J. Wilkerson, K. Gard, A. Schuchinsky, and M. Steer, "Electro-thermal theory of intermodulation distortion in lossy microwave components," *IEEE Trans. on Microwave Theory and Techniques*, vol. 56, no. 12, pp. 2717–2725, Dec. 2008.

[38] J. Wilkerson, K. Gard, and M. Steer, "Automated broadband high-dynamic-range nonlinear distortion measurement system," *IEEE Trans. on Microwave Theory and Techniques*, vol. 58, no. 5, pp. 1273–1282, may 2010.

[39] C. Vicente and H. Hartnagel, "Passive-intermodulation analysis between rough rectangular waveguide flanges," *IEEE Trans. on Microwave Theory and Techniques*, vol. 53, no. 8, pp. 2515–2525, Aug. 2005.

[40] C. Vicente, D. Wolk, H. Hartnagel, and D. Raboso, "An experimental investigation on passive intermodulation at rectangular waveguide interfaces," in *2006 IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 2006, pp. 242–245.

[41] G. Mazzaro, M. Steer, and K. Gard, "Intermodulation distortion in narrowband amplifier circuits," *IET Microwaves, Antennas Propagation*, vol. 4, no. 9, pp. 1149–1156, Sep. 2010.

[42] G. Mazzaro, M. Steer, K. Gard, and A. Walker, "Response of RF networks to transient waveforms: Interference in frequency-hopped communications," *IEEE Trans. on Microwave Theory and Techniques*, vol. 56, no. 12, pp. 2808–2814, Dec. 2008.

[43] J. Wetherington and M. Steer, "Robust analog canceller for high-dynamic-range radio frequency measurement," *IEEE Trans. on Microwave Theory and Techniques*, vol. 60, no. 6, pp. 1709–1719, Jun. 2012.

[44] A. Christianson and W. Chappell, "Measurement of ultra low passive intermodulation with ability to separate current/voltage induced nonlinearities," in *2009 IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 2009, pp. 1301–1304.

[45] J. Russer, A. Ramachandran, A. Cangellaris, and P. Russer, "Phenomenological modeling of passive intermodulation (PIM) due to electron tunneling at metallic contacts," in *2006 IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 2006, pp. 1129–1132.

[46] J. Simmons, "Generalized thermal j-v characteristic for the electric tunnel effect," *Applied Physics*, vol. 35, no. 9, pp. 2655–2658, Sep. 1964.

[47] J. Henrie, A. Christianson, and W. Chappell, "Engineered passive nonlinearities for broadband passive intermodulation distortion mitigation," *IEEE Microwave and Wireless Components Letters*, vol. 19, no. 10, pp. 614–616, Oct. 2009.

[48] J. Wetherington and M. Steer, "Standoff acoustic modulation of radio frequency signals in a log-periodic dipole array antenna," *IEEE Antennas and Wireless Propagation Letters*, vol. 11, no. 8, p. 1, Aug. 2012.

[49] A. Hati, C. Nelson, and D. Howe, "Effect of vibration on P and AM noise of oscillatory and non-oscillatory components at 10 GHz," in *2009 IEEE Int. Frequency Control Symp., Joint with the 22nd European Frequency and Time Forum*, Apr. 2009, pp. 524–529.

[50] "E-20-01A: ECCS Standard on Multipaction and Test," http://www.ecss.nl.

4.11 Exercises

1. An amplifier consists of three cascaded stages with the following characteristics:

	Stage 1	Stage 2	Stage 3
Gain (dB)	−3	15	5
NF (dB)	3	2	2

   (a) What is the overall gain of the amplifier?
   (b) What is the overall noise figure of the amplifier?

2. Briefly describe the effect of a lossy filter on SNR. Consider signals at the input and output of the filter.

3. What is the available noise power of a 50 Ω resistor in a 10 MHz bandwidth. The resistor is at standard temperature.

4. A 50 Ω resistor a 20 Ω resistor are in shunt. If both resistors have a temperature of 300 K, what is the total available noise power spectral density of the shunt resistors?

5. The thermal noise power at the output of a system is 1 fW and the shot noise power is 1 fW. What is the available white noise power?

6. A 2 GHz amplifier in a 50 Ω system has a bandwidth of 10 MHz, a gain of 40 dB, and a noise figure of 3 dB. The amplifier is driven by a circuit with a Thevenin equivalent resistance of 50 Ω held at 290 K (standard temperature). What is the available noise power at the output of the amplifier?

7. A 30 dB attenuator is terminated at Port 2 in a matched resistor and both are at 290 K. What is the noise temperature at Port 1 of the attenuator?

8. A 20 dB attenuator is terminated in a matched resistor and both are held at 30°C. What is the noise temperature at the input of the attenuator in kelvin?

9. The effective noise temperature at the coaxial output of an antenna is 100 K. The antenna is connected to a bandpass filter with a bandwidth of 20 MHz and an insertion loss of 1 dB. [Parallels Examples 4.2 and 4.3]
   (a) What is the available noise power in a 20 MHz bandwidth at the output of the antenna?
   (b) What is the noise figure of the bandpass filter (consider only the passband)?
   (c) What is the excess noise power at the output of the filter? (Consider only the passband).
   (d) What is the total available noise power in the passband at the output of the filter?

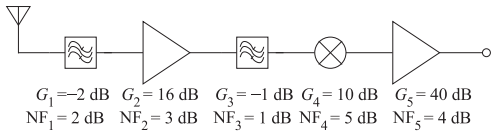
10. A receive amplifier with a gain of 30 dB, a noise figure of 2 dB, and bandwidth of 5 MHz is connected to an antenna which has a noise temperature of 20 K. [Parallels Example 4.2]
   (a) What is the available noise power presented to the input of the amplifier in the 5 MHz bandwidth (recall that the antenna noise temperature is 20 K?
   (b) If instead the input of the amplifier is connected to a resistor held at standard temperature, what is the available noise power presented to the input of the amplifier in the 5 MHz bandwidth?
   (c) What is the noise factor of the amplifier?
   (d) What is the excess noise power of the amplifier referred to the its output?
   (e) What is the effective noise temperature of the amplifier when the amplifier is connected to the antenna with a noise temperature of 20 K. That is, what is the effective noise temperature of the resistor in the Thevenin equivalent circuit of the amplifier output?

11. A receive amplifier has a bandwidth of 5 MHz, a 1 dB noise figure, a linear gain of 20 dB. The minimum acceptable SNR is 10 dB.
   (a) What is the output noise power in dBm?
   (b) What is the minimum detectable output signal in dBm?
   (c) What is the minimum detectable input signal in dBm?

12. A 75 Ω attenuator has a loss of 16 dB and is between a source with a Thevenin impedance of 75 Ω and a load of 75 Ω.
   (a) What is the noise power, N_i, available from the 75 Ω source resistor at standard temperature (290 K) in a 1 MHz bandwidth?
   (b) Now consider that the source is connected to the attenuator which is also connected to the load. If the source generates a modulated signal that is 1 MHz wide and has an available power, S_i, of 10 fW, what is SNR_i at the input to the attenuator at standard temperature?
   (c) With the attenuator connected to the source, what is the Thevenin equivalent impedance looking into the output of the attenuator?
   (d) Calculate the noise power, N_o, available from the attenuator with the source attached at standard temperature (290 K) in a 1 MHz bandwidth?
   (e) What is the signal power, S_o, delivered to

the load?
(f) What is the SNR at the load, SNR_o?
(g) What is the noise factor of the attenuator?
(h) What is the noise figure of the attenuator?

13. The system shown below is a receiver with bandpass filters, amplifiers, and a mixer. [Parallels Example 4.4]



(a) What is the total gain of the system?
(b) What is the noise factor of the first filter?
(c) What is the system noise factor?
(d) What is the system noise figure?

14. An amplifier consists of three cascaded stages with the following characteristics:

	Stage 1	Stage 2	Stage 3
Gain	10 dB	15 dB	30 dB
NF	0.8 dB	2 dB	2 dB

What is the noise figure (NF) and gain of the cascade amplifier?

15. The front end of a receiver for a cellular phone has a bandpass filter with a 25 MHz passband and a loss in the passband of 2 dB and is followed by two amplifier stages. The first stage has a gain of 20 dB and a noise figure of 0.5 dB and the second stage has a gain of 60 dB and a noise figure of 2 dB.
(a) Sketch the system block diagram.
(b) What is the gain of the system?
(c) What is the noise figure of the filter?
(d) What is the noise figure of the system?
(e) The system is now connected to an antenna with an effective noise temperature of 30 K that delivers a signal of 10 pW to the bandpass filter. Determine the noise temperature at the output of the system and hence the output noise power in the 25 MHz bandwidth. First calculate the excess noise temperature added by the system to the output.
1. Determine the signal-to-noise ratio at the output of the front-end system.

16. The RF front end of a communications unit consists of an amplifier followed by a mixer. The amplifier has a gain of 20 dB and a noise figure of 4 dB. The mixer has a conversion loss of 6 dB and a double-sideband noise figure of 8 dB.
(a) Why is the double-sideband noise figure sometimes used with a mixer but not with an amplifier?
(b) If the amplifier has sufficient bandwidth to amplify both the RF and image frequency, what is the noise figure of the cascade? Note that the overall noise figure is a single-sideband noise figure.

17. The first stage of a two-stage amplifier has a linear gain of 40 dB and a noise figure if 3 dB. The second stage has a gain of 10 dB and a noise figure of 5 dB.
(a) What is the overall gain of the amplifier?
(b) What is the overall noise figure of the amplifier?

18. A subsystem consists of a matched filter with an insertion loss of 2 dB then an amplifier with a gain of 20 dB and a noise figure, NF, of 3 dB.
(a) What is the overall gain of the subsystem?
(b) What is NF of the filter?
(c) What is NF of the subsystem?

19. A subsystem consists of a matched amplifier with a gain of 20 dB and a noise figure of 2 dB, followed by a 2 dB attenuator, and then another amplifier with a gain of 10 dB and NF of 3 dB.
(a) What is the overall gain of the subsystem?
(b) What is NF of the attenuator?
(c) What is NF of the subsystem?

20. Consider a digitally modulated signal and briefly describe the impact of a nonlinear amplifier on the signal. You must include several negative effects. Use one or more diagrams.

21. An amplifier has a linear gain of 30 dB and an output-referred 1 dB gain compression point of 13 dBm. What is the input-referred 1 dB gain compression point of the amplifier?

22. An amplifier has a linear gain of 30 dB and an input-referred 1 dB gain compression point of -30 dBm. What is the output-referred 1 dB gain compression point of the amplifier?

23. An amplifier has an output power of 10 dBm when the gain of a single tone is compressed by 1 dB. What is the maximum output power of an undistorted 64-QAM signal? (A 64-QAM signal which has a PMEPR of 7.8 dB. [Parallels Example 4.6]

24. The input-referred 1 dB gain compression point of an amplifier with a linear gain of 30 dB is 0 dBm.
(a) What is the gain of the amplifier at 1 dB gain compression?
(b) What is the output power at 1 dB gain compression?
(c) Consider amplifying an 8-PSK signal with a PMEPR of 3.3 dB. What is the maximum output power of the undistorted 8-PSK signal?

Maximum acceptable distortion is when the envelope peak is compressed by 1 dB gain.

25. The gain of an amplifier at the 1-dB gain compression point is 40 dB and the input power is -7 dBm.
    (a) What is the power of the amplifier's output signal?
    (b) If the power input to the amplifier is reduced to -20 dBm, what is the amplifier's output power now?
26. Briefly describe intermodulation distortion with a a two-tone signal. Use a diagram.
27. Briefly describe what is meant by 1 dB gain compression. Use a diagram.
29. A single-stage amplifier has a linear gain of 16 dB, an output 1 dB gain compression point of 10 dBm, and an output-referred third-order intercept point OIP3 = 30 dBm. The noise floor at the output of the amplifier is -60 dBm. The communication protocol has a minimum SNR, $\mathrm{SNR}_{\mathrm{MIN}}$, of 6 dB.
    (a) What is the dynamic range of the amplifier?
    (b) What is the SFDR of the amplifier?
30. A receiver system comprising a filter and two cascaded amplifiers has an overall linear gain of 80 dB, an output 1 dB gain compression point of -10 dBm, and an output-referred third-order intercept point, OIP3 = 10 dBm. The noise floor at the output of the amplifier is -80 dBm. What is the spurious free dynamic range of the receiver?
31. A power amplifier has a linear gain of 20 dB, an output 1 dB gain compression point of 30 dBm, and an output-referred third-order intercept point OIP3 = 60 dBm. The noise floor at the output of the amplifier is -70 dBm. What is the dynamic range of the amplifier if the required minimum SNR at the output is 6 dB?
32. A room-temperature two-stage amplifier in a receiver has a bandwidth of 100 MHz, a noise figure of 3 dB, a linear gain of 32 dB, and an output-referred third-order intercept point, OIP3, of 27 dBm. The minimum SNR of the receiver system is 16 dB.
    (a) What is the output noise power in dBm?
    (b) What is the difference between the input- and output-referred spurious free dynamic ranges?
    (c) What is the SFDR in dB?
    (d) What is the difference between the input- and output-referred dynamic ranges?
    (e) What is the minimum detectable output signal in dBm?
    (f) What is the output-referred DR in dB?
33. When determining the dynamic range of an amplifier the gain compression level is not used. Briefly discuss why.

4.11.1 Exercises by Section

†challenging, ‡very challenging

§4.2 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11
§4.3 12†, 13, 14, 15†, 16†, 17, 18, 19
§4.5 20, 21, 22, 23, 24, 25, 26, 27, 28
§4.6 29, 30, 31, 32, 33

4.11.2 Answers to Selected Exercises

19(b) 60 dB
13(d) 5.17 dB
12(g) 39.8
17(b) 3 dB

CHAPTER 5

Passive Modules

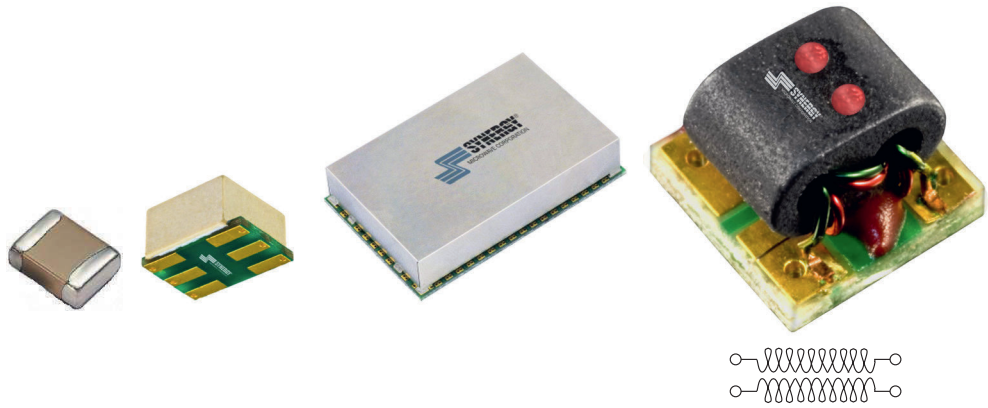
5.1 Introduction to Passive Modules 165
5.2 Diodes 167
5.3 Case Study: Tunable Resonator with Varactor Diode Stack 170
5.4 Switch 172
5.5 Ferrite Components: Circulators and Isolators 177
5.6 Summary 181
5.7 References 182
5.8 Exercises 182

5.1 Introduction to Passive Modules

Most RF systems are composed of a cascade of modules each of which is separately designed and characterized. They usually have matched $50\ \Omega$ input and output impedances so that modules can be freely interconnected. In high volume applications, several modules could be monolithically integrated, but even then, design is based on the concept of modules. Many modules are available "off the shelf" and high-performance RF systems can be constructed using commercially available modules, see Figure 5-1.

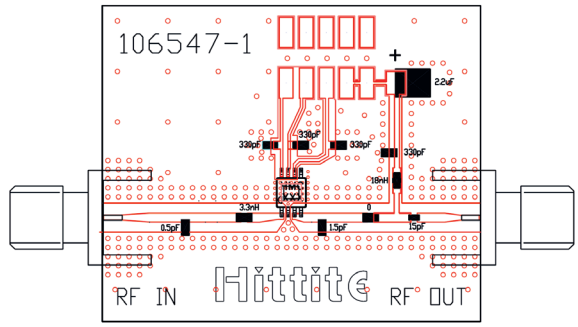
Many modules are available mounted on evaluation boards complete with biasing components and SMA connectors so that RF modules can be interconnected using coaxial cables. For example, an evaluation board is shown in Figure 5-2 for a power amplifier. The engineer needs to attach supply, V_s, ground, and control voltage V_{ctl}, and the evaluation board provides a working amplifier. Most RF and microwave engineers work at the circuit board level and begin system design using modules. Some companies develop some of their own proprietary modules, thus providing a competitive advantage, but still use many modules developed by others. An estimate sometimes made is that the value of RF and microwave circuit-board-based module-using engineering is 30 times larger than the value of module and RF integrated circuit engineering.

The modules that comprise a receiver are shown in Figure 5-3. Beginning with the bandpass filter after the antenna, each module contributes noise and nonlinear distortion. The system design objectives are generally to maximize dynamic range, the region between the signal being sufficiently above the noise level to be detected but before nonlinear distortion introduces spurious

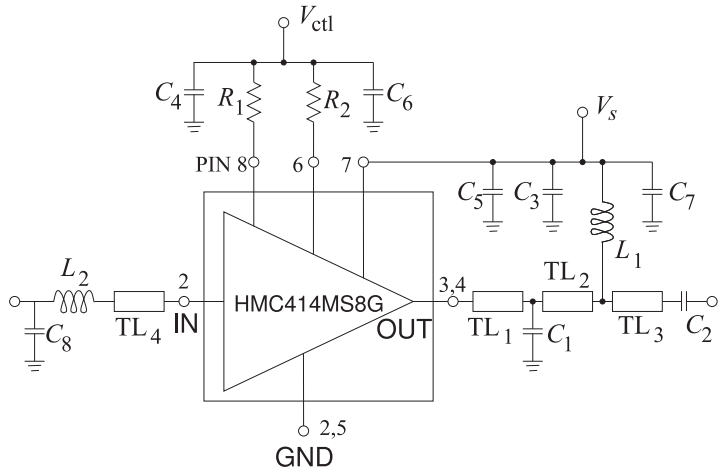


(a) Resistor (b) Mixer (c) Synthesizer (d) Transformer

Figure 5-1: Modules in surface-mount packages. Copyright Synergy Microwave Corporation, used with permission [1].



(a) Evaluation board



(b) Schematic

Figure 5-2: Evaluation board for the HMC414MS8G GaAs InGaP HBT MMIC power amplifier module operating between 2.2 and 2.8 GHz. The amplifier provides 20 dB of gain and +30 dBm of saturated power at 32% PAE from a +5V supply voltage. The amplifier can also operate with a 3.6 V supply selectable by the resistors R_1 and R_2. Copyright Hittite Microwave Corporation, used with permission [2].

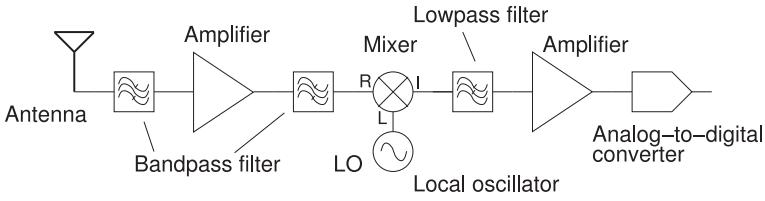


Figure 5-3: Receiver as a cascade of modules.

Component	Symbol
Diode, general (including Schottky)[1]	
IMPATT diode[1]	
Gunn diode	
PIN diode[1,2]	
Light emitting diode (LED)[1]	
Rectifier[1]	
Tunnel diode[1]	
Varactor diode[1]	or
Zener diode[1]	

Table 5-1: IEEE standard symbols for diodes and a rectifier [3]. ([1]In the direction of anode (A) to cathode (K). [2]Use symbol for general diode unless it is essential to show the intrinsic region.)

signals that limit the detectability of signals. At the same time, it is usual to try and minimize the power consumed. This is particularly important in mobile applications.

This chapter introduces many passive modules;modules and microwave functional elements that do not require external supplies with he exception of perhaps control voltages..

5.2 Diodes

Diodes are two-terminal devices that have nonlinear current-voltage characteristics. The most common diodes used in microwave engineering are listed in Table 5-1 along with their standard symbols. Ideally a **rectifier** allows current to flow in one direction and not in the other. A general diode, usually a junction diode or a **Schottky diode**, is one type of rectifier, but rectifiers are more general, and, for example, they can be realized using vacuum devices.

Junction and Schottky Diodes

Junction diodes are two-terminal devices that derive their characteristics from the barrier effect that occurs at the junction of two different types of semiconductor (one with excess free holes and one with excess free electrons), or at the interface of a metal and a semiconductor. The result is an asymmetric current-voltage characteristic, as shown in Figure 5-4(a). This is not an ideal rectifier characteristic, as it requires a threshold voltage to be reached before there is appreciable current flow. Semiconductor-based diode

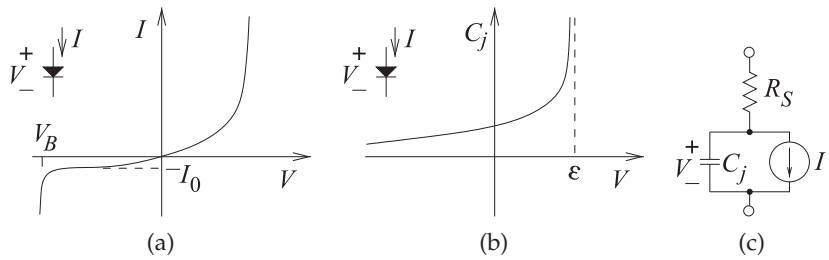


Figure 5-4: Characteristics of a pn junction diode or a Schottky diode: (a) current-voltage characteristic; (b) capacitance-voltage characteristic; and (c) diode model.

characteristics are described by [4–6]

$$I = I_0 \left[\exp\left(\frac{qV}{nkT} \right) - 1 \right], \tag{5.1}$$

where V is the voltage across the junction, $q(= -e)$ is the absolute value of the charge of an electron, k is the Boltzmann constant ($1.37 \cdot 10^{-23}$ J/K), and T is the absolute temperature (in kelvin). I_0 is the reverse saturation current and is small, with values ranging from 1 pA to 1 nA. The quantity n is the diode ideality factor, with $n = 2$ for graded-junction pn junction diodes and $n = 1.0$ for step-junction diodes where the interface between p-type and n-type semiconductor materials is abrupt. The abrupt junction is most closely realized by a Schottky diode, where a metal forms one side of the interface (typically) replacing the p-type semiconductor. Carriers recombine quickly in the metal, much faster than they would in a semiconductor. Hence a Schottky diode operates at higher frequencies than does a pn junction diode. A unity ideality factor, n, of 1 is the best that can be achieved, resulting in the strongest nonlinearity. When the applied voltage is sufficiently positive to cause a large current to flow, the diode is said to be forward biased. When the voltage is negative, the current flow is negligible and the diode is said to be reverse biased. At sufficiently large reverse bias, electrons are ripped from the valence bands of the semiconductor atoms and the current rapidly increases in a process called impact ionization or avalanche. The voltage at which this occurs is called the reverse breakdown voltage.

In a semiconductor diode, charge is separated over distance and so a diode has appreciable capacitance, called the junction capacitance, mathematically modeled as

$$C_j(V) = \frac{C_{j0}}{(1 - (V/\phi))^{\gamma}}, \tag{5.2}$$

where ϕ is the built-in potential difference across the diode. This capacitance profile is shown in Figure 5-4(b). The built-in potential is typically 0.6 V for silicon diodes and 0.75 V for GaAs diodes. The doping profile can be adjusted so that γ can be less than the ideal $\frac{1}{2}$ of an abrupt junction diode.

Current must flow through bulk semiconductor before reaching the active region of the semiconductor diode, and so there will be a resistive voltage drop. Combining effects leads to the equivalent circuit of a pn junction or Schottky diode shown in Figure 5-4(c).

Varactor Diode

A varactor diode is a pn junction diode operated in reverse bias and optimized for good performance as a tunable capacitor. Ideally it has low

reverse saturation current, high reverse breakdown voltage, and a specific capacitance profile designed for a particular application.

A common application of a varactor diode is as the tunable element in a voltage-controlled oscillator (VCO) where the varactor, with voltage-dependent capacitance, C, is part of a resonant circuit (often called a tank circuit) with a lumped inductor, L. The resonant frequency of the tank circuit is proportional to $1/\sqrt{LC}$, so by applying a voltage to a varactor diode, C changes, the resonant frequency of the tank circuit is tuned, and the oscillation frequency is changed. The capacitance versus voltage is described by Equation (5.2) and voltage dependence is shown in Figure 5-4(b).

PIN Diode

A PIN diode is a variation on a pn junction diode with a region of intrinsic semiconductor (the I in PIN) between the p-type and n-type semiconductor regions. The properties of the PIN diode depend on whether there are carriers in the intrinsic region. The PIN diode has the current-voltage characteristics of a pn junction diode at low frequencies; however, at high frequencies it looks like a linear resistor, as carriers in the intrinsic region move slowly. When a forward DC voltage is applied to the PIN diode, the intrinsic region floods with carriers, and at microwave frequencies the PIN diode is then modeled as a low-value resistor. At high frequencies there is not enough time to remove the carriers in the intrinsic region, so even if the total voltage (DC plus RF) across the PIN diode is negative, there are carriers in the intrinsic region throughout the RF cycle. If the DC voltage is negative, carriers are removed from the intrinsic region and the diode looks like a large-value resistor at RF. The PIN diode can be used as a microwave switch controlled by a DC voltage.

Zener Diode

Zener diodes are pn junction or Schottky diodes that have been specially designed to have sharp reverse breakdown characteristics [5]. They can be used to establish a voltage reference or, used as a limiter diode, to provide protection of more sensitive circuitry. As a limiter, they are found in communication devices in a back-to-back configuration to limit the voltages that can be applied to sensitive RF circuitry.

LED Diode

In semiconductor diodes, the recombination of holes and electrons can result in photons being generated. This effect is enhanced to form **light-emitting diodes** (LEDs) [4, 5, 7, 8].

5.3 Case Study: Tunable Resonator with Varactor Diode Stack

Varactors provide a tunable capacitance with the bias voltage across the varactor, a reverse biased diode, changing the capacitance of the diode. The applied bias voltage also changes the series resistance of the diode, see Figure 5-4. The characteristics of a microwave varactor are shown in Figure 5-5. At a reverse bias voltage $V_R = 1$ V the junction capacitance $C_j = 7$ pF and the series resistance is $1.8\ \Omega$. At 5 GHz the diode impedance $Z_D(1\ V) = 1.8 - j4.54\ \Omega$ and its $Q(1\ V) = 2.5$. At a reverse bias voltage $V_R = 10$ V, the junction capacitance i$C_j = 1$ pF, and the series resistance is $0.4\ \Omega$. At 5 GHz the diode impedance $Z_D(1\ V) = 0.4 - j31.8\ \Omega$ and its $Q(1\ V) = 80$.

Generally a design proceeds based on a chosen system impedance which is often $50\ \Omega$ so that the design space includes realizable impedances above and below a system impedance. One way of effectively increasing the impedance of a varactor diode is to stack diodes. This is shown in Figure 5-6 where four

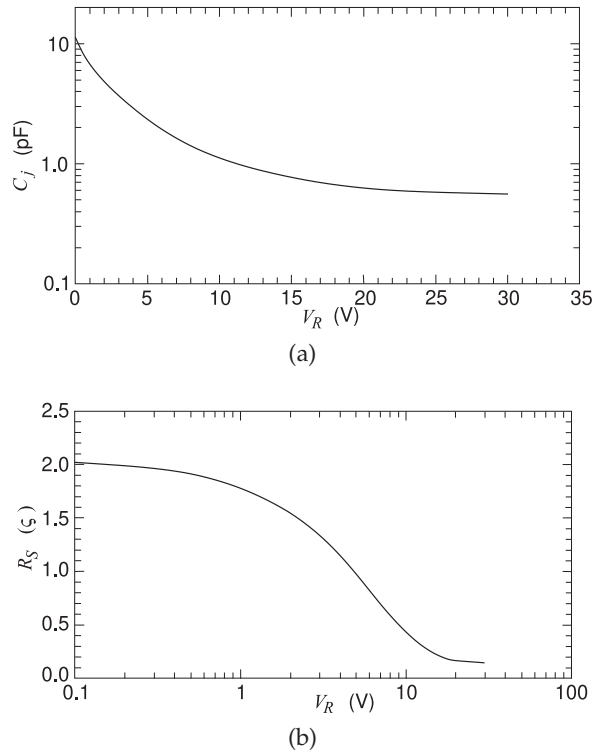


Figure 5-5: Characteristics of a varactor diode (Toshiba model JDV2S71E1, a silicon epitaxial planar diode) as a function of reverse bias voltage V_R: (a) capacitance C_j; and (b) series resistance R_S.

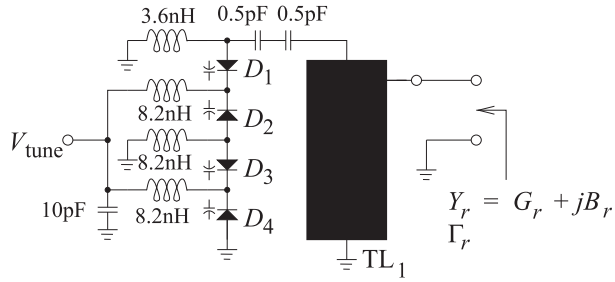


Figure 5-6: A stub-loaded diode stack resonator designed to produce a variable capacitance in the 4–6 GHz range. The element labeled TL_1 is a low impedance microstrip line. The choke inductor, $L_{CHOKE} = 8.2$ nH, presents an RF open circuit and is part of the bias circuit. Each varactor diode (D_1–D_4) is model JDS2S71E.

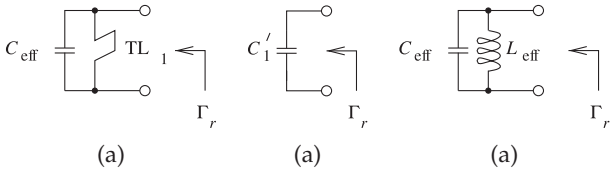


Figure 5-7: Equivalent circuits of the stub-loaded diode stack resonator of Figure 5-6.

varactor diodes $D_1, \ldots, D_4$ are shown together with the biasing arrangement so that the each of the diodes can be adjusted by the single tuning voltage V_{tune} applied through the 3.6 nH and 8.2 nH choke inductors. The choke inductors provide a DC short and an RF open (actually an impedance magnitude of $258\ \Omega$ for the 8.2 nH inductor at 5 GHz). Figure 5-6 is the schematic of a tunable resonator, used between 4 and 5.3 GHz to tune an oscillator, see [9] and Section 5.6 of [10].

The two 0.5 pF capacitors in Figure 5-6 scale the capacitance of the diode stack and increase the effective Q of the variable capacitance. The shorted transmission line presents an inductance impedance in shunt with the augmented diode stack. A lumped inductor is not used here as it would have too high a loss. The equivalent circuit of the resonator is shown in Figure 5-7. The resonator network is designed to be resonant at a frequency above the operating frequency. At the oscillation frequency the resonator network must present an inductance to resonate with the capacitance of the active device in an oscillator. In addition, for oscillator stability, the effective inductance must have a particular derivative of admittance with respect to frequency.

The two 0.5pF capacitors between TL_1 and the varactor stack combine to provide an effective capacitance of 0.25 pF which has an impedance of $127\ \Omega$ at 5 GHz. So it is not a DC blocking capacitor which would have a very low impedance at RF, but it does block DC so that V_{tune} is not shorted to ground through the transmission line. The RF function of the two series 0.5 pF capacitors is to reduce the value of the variable capacitance, increase the operating reactance, and increase the Q of the effective capacitance. Each varactor diode has a 0 V capacitance of 11 pF and a capacitance at 9 V of 2.15 pF. With the four varactors in series this becomes 2.75 pF total capacitance at 0 V and 0.54 pF at 9 V. The 0.25 pF capacitance in series reduces this to 0.229 pF at 0 V tuning voltage and 0.171 pF at 9 V tuning voltage for a capacitance tuning ratio of 0.229 pF/0.171 pF = 1.34. The capacitance tuning range has been reduced from the original ratio of 2.75 pF/ 0.54 pf = 5.05. The tuning range of the resonant frequency of a parallel LC resonator is the square root of this, i.e. 1.16 (which comes from the resonant frequency being $1/\sqrt{LC}$). If this is not enough the two 0.5 pF capacitors can be changed to achieve the required overall capacitance tuning ratio.

The 10 pF capacitor between the V_{tune} terminal and ground is there to ensure an RF short circuit. The magnitude if the impedance of this capacitor at 5 GHz is $3.2\ \Omega$, a factor of ten away from the $50\ \Omega$ system impedance. While at RF the terminal at V_{tune} is ideally a short circuit to ground, there will always be some wiring inductance.

The characteristics of the resonator can be understood by looking at the plot of the simulated resonator reflection coefficient, Γ_r, on a Smith chart, see Figure 5-8. Γ_r is shown for two tuning voltages as well as the device locus of the inverse of the input reflection coefficient of a transistor circuit

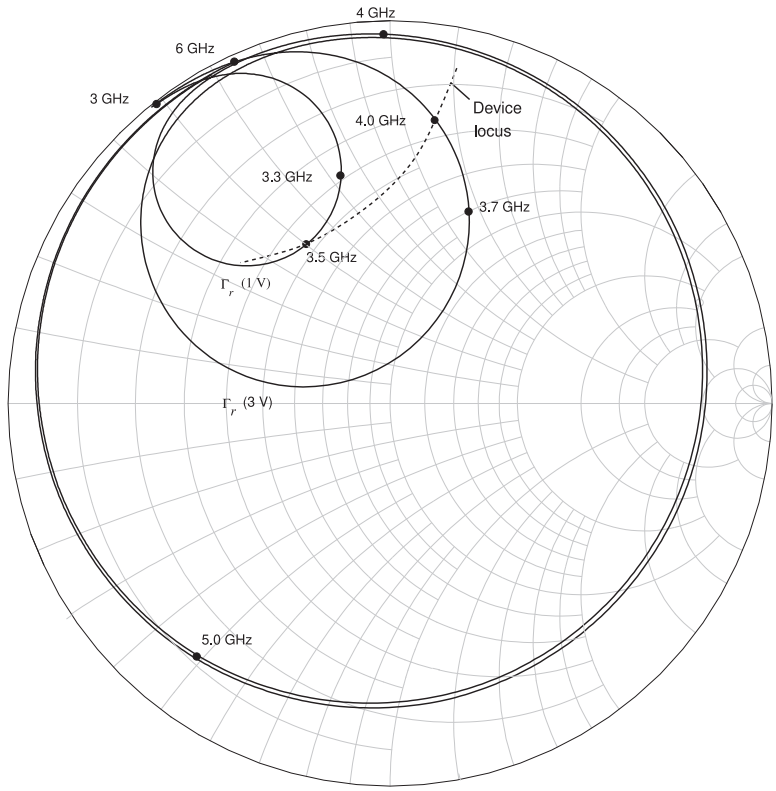


Figure 5-8: Simulated reflection coefficient of the resonator, Γ_r on a 50 Ω Smith chart for two tuning voltages (1 V and 3 V). The drive locus shown is the inverse of the input reflection coefficient of an active network that would be connected at the same port as for Γ_r. The intersection of Γ_r and the drive locus determines the operating frequency.

connected to the resonator. The combination of the resonator and the active device forms an oscillator. The intersection of the active device and resonator loci selects the oscillation frequency of the oscillator. The oscillator presents a capacitance to the resonator and so the resonator must present an inductance to the active device. The frequency of the oscillator is varied by changing Γ_r of the resonator.

A photograph of the resonator is shown in Figure 5-9. The measured Γ_r is shown in Figure 5-10 and this corresponds to the simulated plot in Figure 5-8. The two plots are roughly in agreement but there is a discrepancy because it is very hard to do these measurements at 5 GHz, and the models of the device components are not exact. However the main reason for the discrepancy is that this is a measurement of a resonant circuit and small discontinuities can have a dramatic effect. What is presented here is typical of microwave circuit design and why tremendous intuition is required to design at microwave frequencies and to interpret measurements.

5.4 Switch

Microwave switches are commonly used to alternately connect an antenna to a transmitter or a receiver. In some communication systems, such as GSM and some modes of 4G and 5G, a phone does not transmit and receive simultaneously. Consequently a switch can be used to separate the transmitted and received signals. In multiband phones, a switch is used to connect the correct transmitter and receiver, which are band specific, to the antenna. In radar systems, switches are used to steer an antenna beam by changing the phase of the microwave signal delivered to each antenna in an array of antennas. An ideal microwave switch is shown in

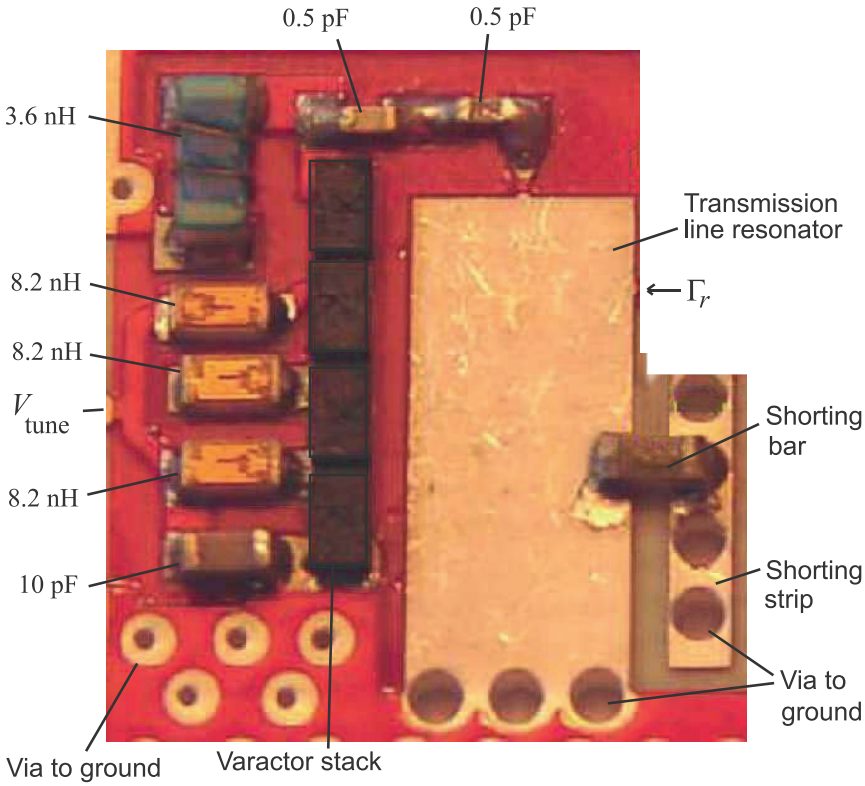


Figure 5-9: Photograph of the microstrip resonator network with surface mount components. The shorting bar is a 0 Ω surface mount resistor and is used to adjust the effective length of the transmission line resonator, TL_1.

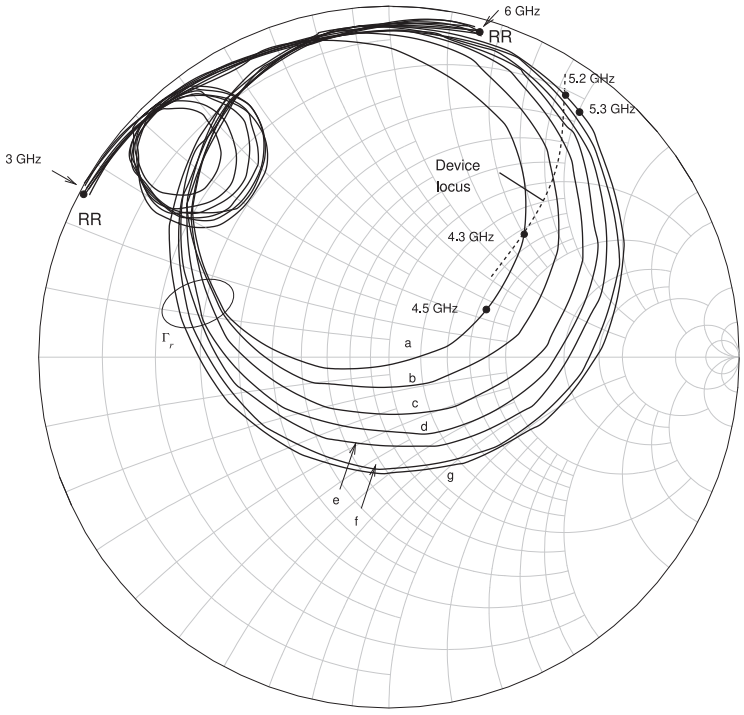


Figure 5-10: Measured reflection coefficient of the resonator network, Γ_r. Curve a, is the characteristic of the resonator network for a tuning voltage, V_{tune}, of 0 V. Curve g is for $V_{\text{tune}} = 9$ V. V_{tune} is equally spaced for Curves a–g. The inverse of the reflection coefficient of an active device connected atthe Γ_r point is shown and indicates that the active device oresents a negative conductance. The intersection of the device locus and Γ_r determines the oscillation frequency of an oscillator. Thus adjusting V_{tune} changes the oscilaltion frequency.

Figure 5-11(a), where an input port, $\mathrm{RF_{IN}}$, and an output port, $\mathrm{RF_{OUT}}$, are shown. For maximum power transfer between the ports the switch should have little loss and thus have low on resistance. At microwave frequencies, realistic switches must be modeled with parasitics and with finite on and off resistances. A realistic model applicable to many switch types is shown in Figure 5-11(b). The capacitive parasitics, the C_Ps, limit the frequency of operation of the switches and the on resistance, R_{ON}, impacts the switch loss. Ideally the off resistance, R_{OFF}, is very large, however, the parasitic shunt capacitance, C_{OFF}, is nearly always more significant. The result is that at high frequencies there is an alternative capacitive connection between the input and output through C_{OFF}. The on resistance of the switch introduces voltage division that can be seen by comparing the ideal connection shown in Figure 5-11(c) and the more realistic connection shown in Figure 5-11(d). From the voltage division ratio, the loss of the switch can be calculated.

Switches are configured to provide connections from one or more inputs to one or more outputs. The configuration of a switch is indicated by poles and throws, and several configurations are shown in Figure 5-12. In microwave applications, single-pole switches are most commonly used and the input is connected to an antenna. For example, the throws would be connected to different bands of a multiband phone .

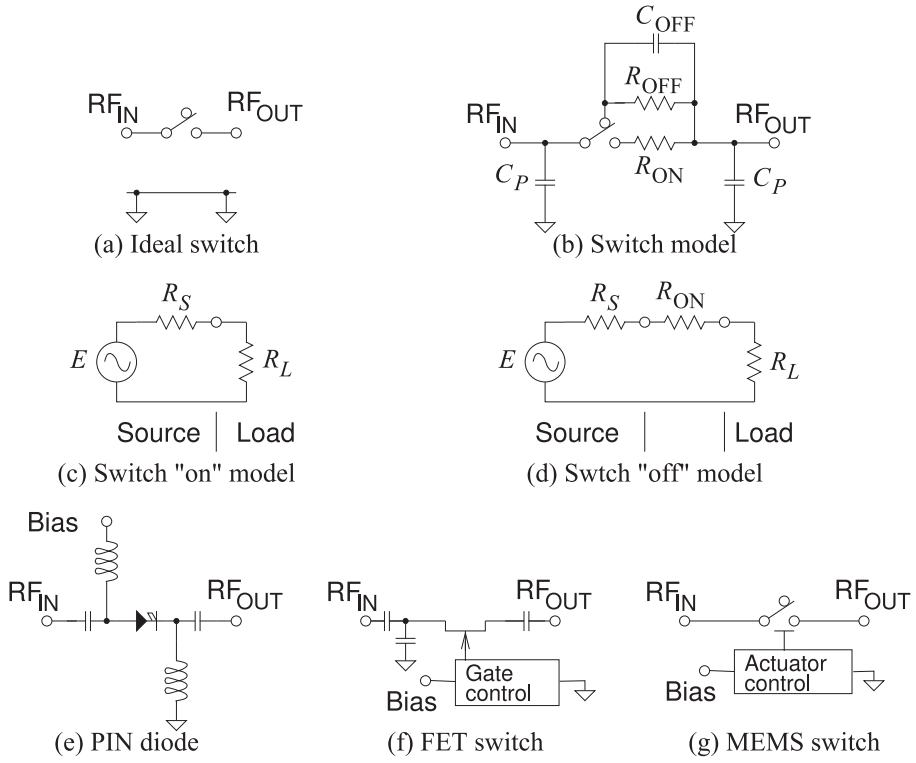


Figure 5-11: Microwave switches: (a) ideal switch connecting $\mathrm{RF_{IN}}$ and $\mathrm{RF_{OUT}}$ ports; (b) model of a microwave switch; (c) ideal circuit model with switch on and with source and load; (d) realistic low-frequency circuit model with the switch on; (e) switch realized using a PIN diode; (f) switch realized using an FET; and (g) switch realized using a MEMS switch.

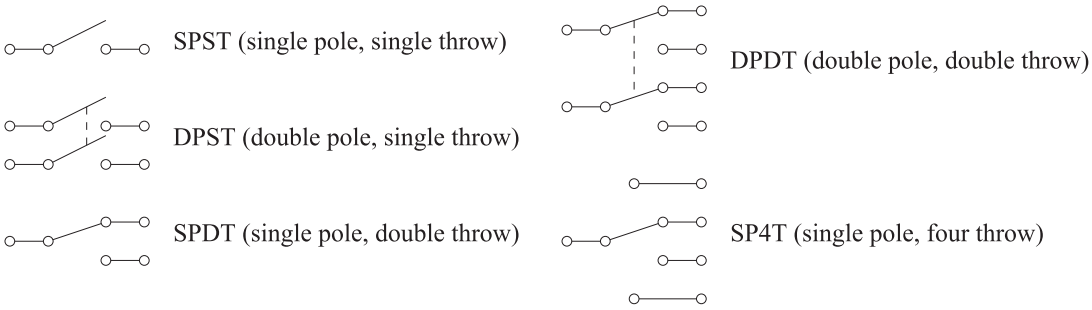

SPST (single pole, single throw)
DPST (double pole, single throw)
SPDT (single pole, double throw)
DPDT (double pole, double throw)
SP4T (single pole, four throw)

Figure 5-12: Switch configurations.

Table 5-2: Typical properties of small microwave switches. (Sources: [1]Radant MEMS, [2]RF Micro Devices, [3]Tyco Electronics,and [2]pSemi™ Corporation.)

Switch type	Configuration	Power handling	Maximum insertion loss	Operating frequency	Actuation voltage	Response time
MEMS[1]	SPDT	0.5 W	0.5 dB	to 10 GHz	90 V	10 μs
MEMS[1]	SPST	4 W	0.8 dB	to 35 GHz	110 V	10 μs
FET (pHEMT)[2]	SPDT	10 W	0.3 dB	to 6.5 GHz	5 V	0.5 μs
FET (pHEMT)[2]	SPDT	0.3 W	1.1 dB	to 25 GHz	5 V	0.5 μs
PIN[3]	SPDT	13 W	0.35 dB	to 2 GHz	12 V	0.5 μs
PIN[3]	SPDT	10 W	0.4 dB	to 6 GHz	12 V	0.5 μs
FET (SOI)[4]	SPDT	2 W	2.7 dB	60 GHz	3.6 V	8 ns

There are four main types of microwave switches: mechanical, PIN diode, FET, and **microelectromechanical system** (**MEMS**) switches. Mechanical switches are nearly lossless but tend to be large, relatively expensive, slow to operate, and are mostly used in laboratory settings. The other switches are of most interest for use in systems. The PIN diode, FET, and MEMS switches are shown in Figures 5-11(e–g), respectively. With these technologies, most higher-order switches are based on interconnections of **SPST switches**. The attributes of these switches are summarized in Table 5-2 for switches that are suitable for cell phone applications. PIN diode switches are the most robust, handling the most RF power, and operating at higher frequencies than either FET- or MEMS-based switches. However, this comes at a price. The PIN diode used is similar to a pn junction diode with the addition of an intrinsic layer between the p- and n-type materials. With applied forward bias the diode has low RF series resistance. In reverse bias, the diode RF resistance is large. Forward bias requires DC current and voltage, so control power is consumed when a PIN diode switch is on. The circuit configuration for an SPST PIN diode switch is shown in Figure 5-11(e). Series bias-decoupling capacitors are required at the RF ports.

A FET makes a good electronic switch; with the correct bias applied to the gate, the drain-source connection looks like a small resistance. Changing the bias to the other extreme removes free carriers from the channel between the drain and source, and a large resistance is the result. Both Si and GaAs

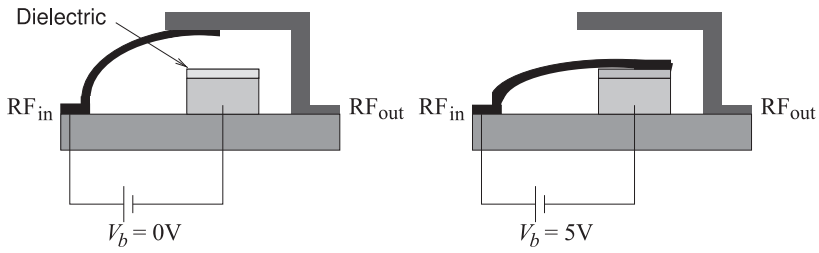


Figure 5-13: RF MEMS switch: (a) the RF_{in} line in contact with the RF_{out} line; and (b) the cantilever beam electrostatically attracted to the pedestal and there is no RF connection.

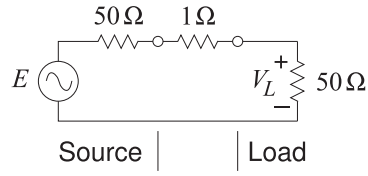


Figure 5-14: Model used in calculating the loss of a switch in a 50 Ω system.

switches are used in cellular handsets, with GaAs switches operating at extended frequencies approaching 6 GHz. The operation of a FET can be described as a variable drain-source resistance with the gate-source voltage controlling the cross section of the channel. The circuit for a FET-based SPST switch is shown in Figure 5-11(f). Series bias-blocking capacitors are required at the RF ports. Control power is only required to change the state of the switch; negligible power is required to maintain the switch state.

A MEMS switch is fabricated using photolithographic techniques similar to those used in semiconductor manufacturing [11–13]. They are essentially miniature mechanical switches with a voltage used to control the position of a shorting arm, which is usually a cantilever or a membrane, see Figure 5-13. As there is no direct connection between the RF signal path and the control circuitry, MEMS switches have inherently high operating frequencies. Power is required to change the switch, but once switching has been accomplished, negligible DC power is required to maintain the connection. Two of the major problems are creating the contact with enough force to reduce the series resistance, and maintaining the structural integrity of the contacts after hundreds of millions of cycles. Generally it is not necessary to make a DC contact, and creating a structure that has a large difference between the on capacitance and off capacitance is sufficient for applications such as tunable matching networks [11].

EXAMPLE 5.1 Insertion Loss of a Switch

What is the insertion loss of a switch with a 1 Ω on resistance when it used in a 50 Ω system?

Solution:

The model to be used for evaluating the insertion loss of the switch is shown in Figure 5-14. The insertion loss is found by first determining the available power from the source and then the actual power delivered to the load. The available input power is calculated by first ignoring the 1 Ω switch resistance. Then there is maximum power transfer from the source to the load. The available input power is

$$P_{Ai} = \frac{1}{2}\frac{\left(\frac{1}{2}E\right)^2}{50} = \frac{E^2}{400}, \tag{5.3}$$

where E is the peak RF voltage at the Thevenin equivalent source generator. The power delivered to the $50\ \Omega$ load is found after first determining the peak load voltage:

$$V_L = \frac{50}{50+1+50}E = \frac{50}{101}E. \tag{5.4}$$

Thus the power delivered to the load is

$$P_D = \frac{1}{2}\frac{V_L^2}{50} = \frac{1}{100}\left(\frac{50}{101}\right)^2 E^2. \tag{5.5}$$

The insertion loss is

$$\text{IL} = \frac{P_{Ai}}{P_D} = \frac{E^2}{400}\frac{100}{E^2}\left(\frac{101}{50}\right)^2 = 1.020\ = 0.086\ \text{dB}. \tag{5.6}$$

5.5 Ferrite Components: Circulators and Isolators

Circulators and isolators are nonreciprocal devices that preferentially route RF signals [14–16]. The essential element of a circulator is a slab or disc of ferrite which, when magnetized, becomes nonreciprocal, with a preferred direction of propagation resulting from what is called the gyromagnetic effect.

5.5.1 Gyromagnetic Effect

An electron has an intrinsic magnetic field described by its magnetic moment $\boldsymbol{m}$ (see Figure 5-15(a)) [17]. This is due to a quantum mechanical property called spin bu the electron does not actually spin. In most materials electron spin occurs in pairs with the magnetic field produced by one electron's spin being canceled by the spin of the other electron. However, in some materials the spin does not occur in pairs and there are regions in which the small magnetic fields produced at the atomic level line up to produce organized regions of directed magnetic field. These regions are called domains, or sometimes **magnetic domains**, and the combined magnetic field from the aligned electron spins is called a magnetic moment.

The most common magnetic material is ferrite, which is a ceramic containing iron(III) oxide (Fe_2O_3). Very strong magnetic effects are obtained with materials containing rare earth elements. The **rare earth** elements, the **lanthanoid elements**, can align the spins of many electrons of each atom and can produce particularly strong magnetic domains and magnetic moments. The most common rare earth elements used in magnetic materials are samarium and neodymium used in samarium-cobalt and neodymium-iron-boron magnets, respectively.

The domains can be aligned by a strong externally generated magnetic

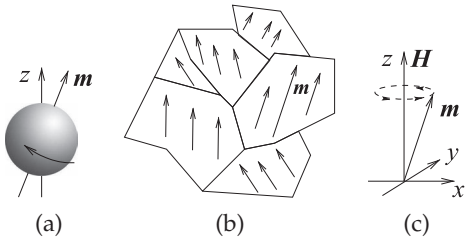


Figure 5-15: Magnetic moments: (a) electron withe magnetic moment $\boldsymbol{m}$; (b), magnetic domains; and (c) progression with a magnetic field bias.

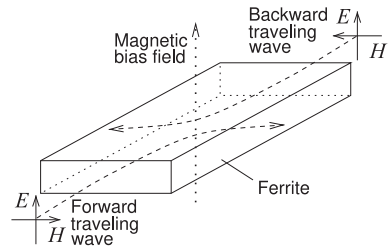


Figure 5-16: Gyromagnetic effect on the propagation of EM waves in a magnetic material (here a ferrite) with an externally applied magnetic bias field.

field (e.g., using an electromagnet), and once the domains are aligned they can stay that way for years even when the external field is removed. This situation is shown in Figure 5-15(b), where the individual domains have been almost completely aligned. Each crystal of the material will usually have many magnetic domains and a domain can grow in size (while another reduces or disappears), but generally a domain will not cross a crystal boundary. The domains can also be partially rotated by a moderate applied magnetic field and when the applied field is removed the domains return to their original alignment. The result is that magnetic energy is stored in much the same way as energy is stored in a spring. The amount of energy that is stored depends on the orientation of the applied magnetic field to the domains. The energy storage capability is described by the permeability of the material. In the absence of a constant biasing magnetic field, the permeability would have three values, one for each of the x, y, and z directions.

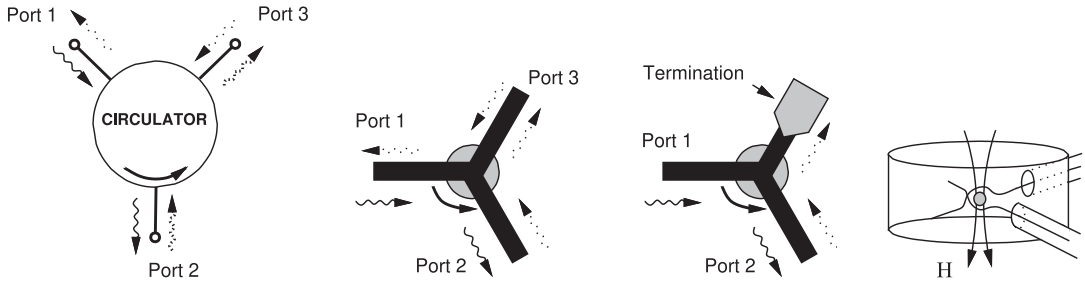
A most interesting microwave property occurs when the magnetic material is biased by a strong DC magnetic field. This situation is depicted in Figure 5-15(c). When a time-varying magnetic field is alsoapplied (e.g., the magnetic component of an EM field), the magnetic moment vector will tend to rotate around the DC magnetic field as shown. This is called the gyromagnetic effect. When the frequency of the applied field corresponds to the characteristic frequency of rotation of the magnetic moment then the effect is called **gyromagnetic** resonance and there is sustained low loss rotation.

Even without resonance (because the magnetic bias field is too small) the gyromagnetic effect affects the way an RF field propagates and this is described by a nine element permeability called a **tensor** (or a **dyadic** or a dyadic tensor) that relates each of the three H field components (in the x-, y-, and z-directions) to each of the three B field components. The permeability of a magnetically biased magnetic material is:

$$[\mu] = \begin{bmatrix} \mu_{xx} & \mu_{xy} & \mu_{xz} \\ \mu_{yx} & \mu_{yy} & \mu_{yz} \\ \mu_{zx} & \mu_{zy} & \mu_{zz} \end{bmatrix} = \begin{bmatrix} \mu_0 & 0 & 0 \\ 0 & \mu & \jmath\kappa \\ 0 & -\jmath\kappa & \mu \end{bmatrix}. \tag{5.7}$$

This tensor can take other forms depending on the orientation of the time-varying magnetic field to the DC biasing magnetic field.

The effect on propagation of an EM field is shown in Figure 5-16. When the EM field is in the magnetized magnetic material, the wave does not travel in a straight line and instead curves, in this case, to the right. Thus forward- and backward-traveling waves diverge from each other and propagation is not reciprocal. This can be used to separate forward- and backward-traveling waves. This is called field displacement.



(a) Circulator schematic (b) Microstrip circulator (c) Microstrip isolator (d) YIG-tuned BPF

Figure 5-17: Ferrite components [21].

5.5.2 Circulator

A circulator exploits the gyromagnetic effect through field displacement [18–20]. The schematic of a circulator is shown in Figure 5-17(a), where the arrows indicate that the signal that enters Port 1 of the circulator leaves the circulator at Port 2 and not at Port 3. Similarly power that enters at Port 2 is routed to Port 3, and power entering at Port 3 is routed to Port 1. In terms of S parameters, an ideal circulator has the scattering matrix

$$\mathbf{S} = \begin{bmatrix} 0 & 0 & S_{13} \\ S_{21} & 0 & 0 \\ 0 & S_{32} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{5.8}$$

A microstrip circulator is shown in Figure 5-17(b), where a disc of magnetized ferrite can be placed on top of a microstrip Y junction to realize a preferential direction of propagation of the EM fields [22]. Design requires choice of the size of the ferrite disc and design of the appropriate magnetic biasing field. In the absence of the biasing magnetic field, the circulation function does not occur.

In addition to the insertion and return losses, the performance of a circulator is described by its isolation, which is its insertion loss in the undesired direction.

5.5.3 Circulator Isolation

The isolation of a circulator is the insertion loss from what is the output port to the input port, i.e. in the reverse direction. Referring to the circulator in Figure 5-17(a), if port 1 is the input port there are two output ports and so there are two isolations equal to the return loss from port 3 to port 2, and from port 2 to port 1. The smaller of these is the isolation quoted if only one value is given. If this was an ideal circulator and port 2 is perfectly matched, then the isolation would be infinite. If port 2 is not perfectly matched then there will be finite isolation from port 3 to port 1. However the most common source of limited isolation is when the circulator is not perfectly matched. Consider that the reflection coefficient looking into each of the circulator's ports is Γ and if the circulator is well designed $\Gamma \ll 1$. Then the S parameters

of the circulator are

$$\mathbf{S} = \begin{bmatrix} \Gamma & \alpha & T \\ T & \Gamma & \alpha \\ \alpha & T & \Gamma \end{bmatrix} \tag{5.9}$$

where T is the transmission factor and is close to 1 for a good circulator and the leakage, α, is small. If the circulator is lossless then the unitary conditions, from Equations (2.148) and (2.149) of [23], are

$$|T|^2 + |\alpha| + |\Gamma|^2 = 1 \tag{5.10}$$

$$\Gamma\alpha^* + T\Gamma^* + \alpha T^* = 0 \longrightarrow \alpha^* + T(\Gamma^*/\Gamma) + T^*(\alpha/\Gamma) = 0 \tag{5.11}$$

Since α is small $\alpha\Gamma$ is negligibly small, so Equation (5.11) indicates that $|\alpha| = |\Gamma|$. Then from Equation (5.10), $|T|^2 = 1 - 2|\Gamma|^2$, and the S parameters of the lossless slightly-mismatched circulator can be written

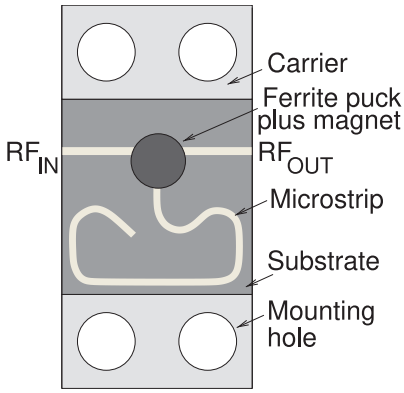
$$\mathbf{S} = \begin{bmatrix} \Gamma & \Gamma & \sqrt{1-2|\Gamma|^2} \\ \sqrt{1-2|\Gamma|^2} & \Gamma & \Gamma \\ \Gamma & \sqrt{1-2|\Gamma|^2} & \Gamma \end{bmatrix}. \tag{5.12}$$

Thus even with ideal external matching at port 2, the isolation is $1/\Gamma$. The quality of the match of a microwave component is typically specified by its VSWR at the ports. Thus the VSWR of a circulator indicates the isolation that can be expected. Actual circulators have a small amount of loss so the isolation and VSWR quoted for a circulator will not conform exactly to the lossless situation considered here.

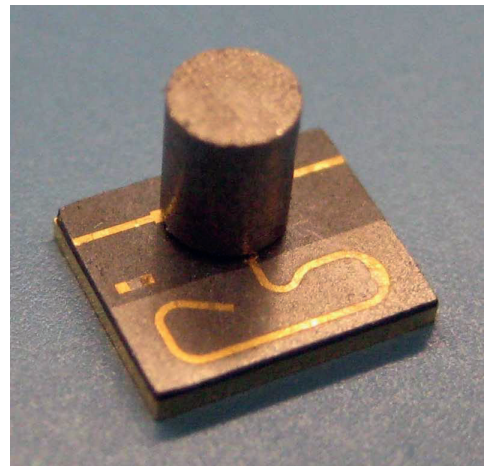
5.5.4 Isolator

Isolators are devices that allow power flow in only one direction. There are two types, one based on field displacement and the other based on gyromagnetic resonance. Ferrite isolators exploiting field displacement are based on a three-port circulator with one of the ports terminated in a matched load. Figure 5-17(c) shows a microstrip isolator based on a three-port circulator. The puck at the center is a magnetic material such as ferrite that when magnetized by a permanent magnet or electromagnet (which is not shown) preferentially supports a counter-clockwise rotating EM wave. So power entering Port 1 as a traveling wave is transferred to the ferrite and emerges at Port 2. Virtually none of the power emerges at Port 3. A traveling wave signal applied at Port 2 appears at Port 3, where it is absorbed in a termination created by resistive material placed on top of the microstrip. The resistive material forms a lossy transmission line and, provided that the lossy line section is long enough, no power is reflected. Thus power can travel from Port 1 to Port 2, but not in the reverse direction. An isolator is commonly used to protect the output of equipment from high reflected signals. A four-port version can implement a duplexer in radar systems and to separate the received and transmitted signals in a transceiver.

A high-frequency microstrip isolator derived from a circulator is shown in Figure 5-18. The input and output lines are at the top and are redirected under the puck. The puck is epoxied to the microstrip substrate and a large biasing magnet is attached on top of the puck. The third port is at the bottom of the figure.



(a) Layout with carrier



(b) Image

Figure 5-18: A microstrip isolator operating from 29 to 31.5 GHz. Isolator in (b) has the dimensions $5 \text{ mm} \times 6$ mm and is 6 mm high. The isolator supports 2 W of forward and reverse power with an isolation of 18 dB and insertion loss of 1 dB. Renaissance 2W9 series, copyright Renaissance Electronics Corporation, used with permission.

Isolators can also exploit gyromagnetic resonance and are then called resonance isolators. A suitable rotation of the RF magnetic field relative to the DC magnetic bias results for one of the directions of propagation in a rectangular waveguide. At the gyromagnetic resonance frequency, RF energy is coupled into the lattice and RF power is absorbed for one direction of propagation and in the other direction the RF signal is little affected.

5.5.5 YIG-Tuned Bandpass Filter

The gyromagnetic effect also enables very sharp variable bandpass filters. A **Yttrium-Iron-Garnet** (**YIG**)-tuned filter is shown in Figure 5-17(d) and the result is that Ports 1 and 2 are only coupled at a precise frequency determined by the state of magnetization of the YIG sphere. With an electromagnet providing a DC magnetic field, the magnetic field can be varied and the resonance frequency, and hence the bandpass frequency, electronically tuned. Microwave spectrum analyzers often use a YIG-tuned filter at the input of the analyzer.

5.6 Summary

The use of modules has become increasingly important in microwave engineering. A wide variety of passive modules were introduced but there are many more. Some microwave modules use unusual effects such as the response of magnetic materials to magnetic fields. Many others use the characteristics of propagating EM fields. The next chapter completes the discussion of modules by covering active modules. A successful microwave designer must have a good knowledge of the types and limitations of available modules.

5.7 References

[1] http://www.synergymwave.com.

[2] http://www.hittite.com.

[3] IEEE Standard 315-1975, Graphic Symbols for Electrical and Electronics Diagrams (Including Reference Designation Letters), Adopted Sept. 1975, Reaffirmed Dec. 1993. Approved by American National Standards Institute, Jan. 1989. Approved adopted for mandatory use, Department of Defense, United States of America, Oct. 1975. Approved by Canadian Standards Institute, Oct. 1975.

[4] B. B. Streetman and S. Banerjee, *Solid State Electronic Devices*, 6th ed. Prentice Hall, 2006.

[5] S. Sze and K. Ng, *Physics of Semiconductor Devices*, 3rd ed. John Wiley & Sons, 2007.

[6] D. Schroder, *Semiconductor Material and Device Characterization*. IEEE Press and Wiley, 2006.

[7] K. Ng, *Complete Guide to Semiconductor Devices*. IEEE Press and Wiley Interscience, 2002.

[8] A. Bergh and P. Dean, "Light-emitting diodes," *Proc. of the IEEE*, vol. 60, no. 2, pp. 156–223, Feb. 1972.

[9] A. Victor and M. Steer, "Reflection coefficient shaping of a 5-GHz voltage-tuned oscillator for improved tuning," *IEEE Trans. on Microwave Theory and Techniques*, vol. 55, no. 12, pp. 2488–2494, Dec. 2007.

[10] M. Steer, *Microwave and RF Design, Amplifiers and Oscillators*, 3rd ed. North Carolina State University, 2019.

[11] L. Katehi and D. Peroulis, "Rf mems components: switches and varactors," in *Multifunctional Adaptive Microwave Circuits and Systems*, M. Steer and W. Palmer, Eds. SciTech Publishing, 2008.

[12] ——, "RF MEMS for reconfigurable circuits and antennas," in *Multifunctional Adaptive Microwave Circuits and Systems*, M. Steer and W. Palmer, Eds. SciTech Publishing, 2008.

[13] G. Rebeiz, *RF MEMS: Theory, Design, and Technology*. Wiley, 2003.

[14] C. Fay and R. Comstock, "Operation of the ferrite junction circulator," *IEEE Trans. on Microwave Theory and Techniques*, vol. 13, no. 1, pp. 15–27, Jan. 1965.

[15] C. K. Queck and L. Davis, "Microstrip and stripline ferrite-coupled-line (fcl) circulator's," *IEEE Trans. on Microwave Theory and Techniques*, vol. 50, no. 12, pp. 2910–2917, Dec. 2002.

[16] A. Borjak and L. Davis, "More compact ferrite circulator junctions with predicted performance," *IEEE Trans. on Microwave Theory and Techniques*, vol. 40, no. 12, pp. 2352–2358, Dec. 1992.

[17] B. Lax and K. J. Button, *Microwave ferrites and ferrimagnetics*. McGraw-Hill, 1962.

[18] D. K. Linkhart, *Microwave circulator design*. Artech House, 2014.

[19] V. G. Harris, "Modern microwave ferrites," *IEEE Trans. on Magnetics*, vol. 48, no. 3, pp. 1075–1104, Mar. 2012.

[20] A. Goldman, *Modern ferrite technology*. Springer Science & Business Media, 2006.

[21] J. Helszajn, *YIG Resonators and Filters*. Wiley, 1985.

[22] U. Milano, J. Saunders, and L. Davis, "A y-junction strip-line circulator," *Microwave Theory and Techniques, IRE Trans. on*, vol. 8, no. 3, pp. 346–351, may 1960.

[23] M. Steer, *Microwave and RF Design, Networks*, 3rd ed. North Carolina State University, 2019.

5.8 Exercises

1. A connector used in a $50\ \Omega$ system introduces a series resistance of $0.5\ \Omega$. What is the insertion loss of the connector?

2. A microwave switch is used in a $75\ \Omega$ system and has a $5\ \Omega$ on resistance. The reactive parasitics of the switch are negligible.

   (a) What is the insertion loss of the switch in the on state?
   (b) What is the return loss of the switch in the on state?

3. A microwave switch is used in a $50\ \Omega$ system and has a $5\ \Omega$ on resistance. The reactive parasitics of the switch are negligible.

   (a) What is the insertion loss of the switch in the on state?
   (b) If the available power of the source is 50 W, what is the power dissipated by the switch?

4. A microwave switch is used at 1 GHz in a $50\ \Omega$ system and it has a $2\ \Omega$ on resistance and a $2\ \text{k}\Omega$ off resistance. The reactive parasitics of the switch are negligible.

   (a) What is the insertion loss of the switch?

(b) What is the isolation of the switch (i.e., what is the insertion loss of the switch when it is in the off state)?

5. A microwave switch is used at 1 GHz in a $50\ \Omega$ system. The switch has a $0.5\ \Omega$ on resistance, a 2 kΩ off resistance, and a parasitic capacitance between the input and output of the switch of 1 pF when the switch is in the off state. Other reactive parasitics of the switch are negligible.

   (a) What is the insertion loss of the switch?
   (b) What is the isolation of the switch ignoring the capacitance of the switch in the off state (i.e., what is the insertion loss of the switch when it is in the off state)?
   (c) What is the isolation of the switch with the off capacitance included?

6. Consider a microwave switch used in a $50\ \Omega$ system. The switch has a $0.5\ \Omega$ on resistance, a 1 kΩ off resistance, and a 0.1 pF off capacitance. Other reactive parasitics of the switch are negligible.

   (a) What is the insertion loss of the switch at 1 GHz?
   (b) What is the isolation of the switch at 1 GHz (i.e., what is the insertion loss of the switch when it is in the off state)?
   (c) What is the insertion loss at 10 GHz?
   (d) What is the isolation at 10 GHz?
   (e) What is the return loss of the switch at 1 GHz when the switch is in the on state?

7. A microwave switch has a $5\ \Omega$ on resistance, and an off resistance of $500\ \Omega$. The reactive parasitics of the switch are negligible.

   (a) What is the insertion loss of the switch when it is used in a $50\ \Omega$ system?
   (b) What is the insertion loss of the switch when it is used in a $100\ \Omega$ system?
   (c) What is the isolation of the switch when it is used in a $50\ \Omega$ system? (That is, what is the insertion loss of the switch when it is in the off state?)
   (d) What is the isolation of the switch when it is used in a $100\ \Omega$ system?

8. The RF front end of a communications unit consists of a switch, then an amplifier, and then a mixer. The switch has a loss of 0.5 dB, the amplifier has a gain of 20 dB, and the mixer has a conversion gain of 3 dB. What is the overall gain of the cascade?

9. Using a schematic, show how an isolator can be obtained from a circulator.

10. A three-port circulator has the S parameters

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

    If port 3 is terminated in a matched load to create a two-port network
    (a) What is the name given to this network?
    (b) Write down the S parameters of the two-port?

11. A three-port circulator has the S parameters

$$\begin{bmatrix} 0 & 0 & 0.5 \\ 20.5 & 0 & 0 \\ 0 & 0.5 & 0 \end{bmatrix}.$$

    Port 3 is terminated in a matched load creating a two-port network.
    (a) Find the S parameters of the two-port.
    (b) What is the return loss in dB at Port 1 if Port 2 is terminated in a matched load?
    (c) What is the insertion loss in dB for a signal applied at Port 1 and leaving at Port 2 with matched source and load impedances?
    (d) What is the insertion loss in dB for a signal applied at Port 2 and leaving at Port 1 with matched source and load impedances?

12. A three-port circulator in a 75-Ω system has the S parameters $\begin{bmatrix} 0.1 & 0 & 0.707 \\ 0.707 & 0.1 & 0 \\ 0 & 0.707 & 0.1 \end{bmatrix}$.

    Port 3 is terminated in 75-Ω creating a two-port network.
    (a) Find the S parameters of the two-port.
    (b) What is the return loss in dB at Port 1 if Port 2 is terminated in 75-Ω?
    (c) What is the insertion loss in dB for a signal applied at Port 1 and leaving at Port 2 with 75-Ω source and load impedances?
    (d) What is the insertion loss in dB for a signal applied at Port 2 and leaving at Port 1 with 75-Ω source and load impedances?

13. A three-port circulator has the S parameters

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

    (a) In no more than 50 words describe the function of a circulator?
    (b) Describe power flow for the circulator with the given S parameters. When power enters one of ports, at which port does power leave? Consider power entering Port 1, then Port 2, then Port 3.

14. An ideal three-port circulator in a $50\ \Omega$ system is configured as an isolator from Port 1 to Port 2. The termination at the third port of the circulator has a value of $45\ \Omega$. What is the isolation of the isolator? (This is the insertion loss in the reverse direction.) [Hint: Use signal flow graph analysis.]

15. Two isolators are used in cascade. Each isolator has an isolation of 20 dB. The isolators are matched so that their input and output reflection coefficients are zero. Determine the isolation of the cascaded isolator system?

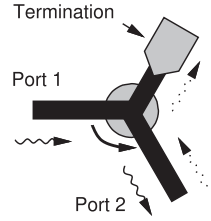
16. A three-port circulator in a 50-Ω system has the S parameters

$$\begin{bmatrix} 0.1 & 0.01 & 0.5 \\ 0.5 & 0.1 & 0.01 \\ 0.01 & 0.5 & 0.1 \end{bmatrix}.$$

If port 3 is terminated in a matched load to create a two-port network

(a) Find the S parameters of the two-port.
(b) What is the return loss in dB at Port 1 if Port 2 is terminated in 50-Ω?
(c) What is the insertion loss in dB for a signal applied at Port 2 and leaving at Port 1 with 50-Ω source and load impedances?
(d) What is the insertion loss in dB for a signal applied at Port 1 and leaving at Port 2 with 50-Ω source and load impedances?
(e) What is is the name of this network?

17. Write down the two-port S parameters of the microstrip isolator shown below.



5.8.1 Exercises By Section

†challenging, ‡very challenging

§5.4 $1, 2^{\dagger}, 3^{\dagger}, 4^{\dagger}, 5^{\dagger}, 6^{\dagger}, 7^{\dagger}$ §5.5 $10, 11, 12, 13, 14^{\dagger}, 15, 16, 17$

5.8.2 Answers to Selected Exercises

3(a) 0.424 dB
2(b) 29.8 dB
5(c) 5.76 dB
7 0.424 dB
8 22.5 dB
15 40 dB

CHAPTER 6

Mixer and Source Modules

6.1 Introduction 185
6.2 Mixer 186
6.3 Single-Ended, Balanced, and Double Balanced Reject Mixers . 197
6.4 Local Oscillator 202
6.5 Voltage-Controlled Oscillator 204
6.6 Phase Detector 206
6.7 Frequency Multiplier 207
6.8 Frequency Divider 208
6.9 Phase Locked Loop 209
6.10 Direct Digital Synthesizer 213
6.11 Diode and Vacuum Sources 214
6.12 Summary 217
6.13 References 217
6.14 Exercises 220

6.1 Introduction

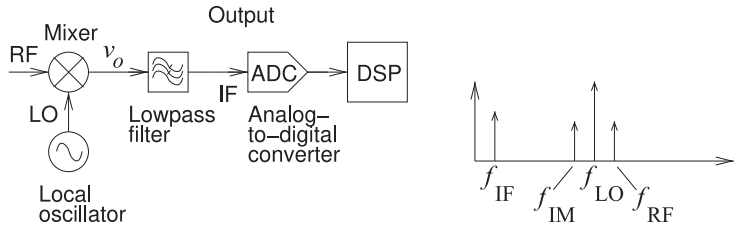
This chapter considers frequency generation and frequency translation modules. In a receiver the basic frequency translation module is a mixer that is used in a receiver to take a signal at a high frequency to a low frequency, the baseband signal, where it can be more easily sampled by an analog-to-digital converter (ADC) and then digitally processed. In a transmitter the signal flow is reversed and a low frequency baseband signal is translated to a high-frequency signal where it can be more easily transmitted. A local oscillator (LO) is input to the mixer to provide the translation mechanism by varying the impedance of an element at the LO frequency. A voltage-controlled oscillator (VCO) allows for agility of the oscillator. A VCO with a phase detector embedded in a phase-locked loop (PLL) provides a well-defined variable frequency oscillation signal. Special RF signal sources are based either on special characteristics of certain semiconductor devices or on vacuum devices in which a beam of electrons interacts with an RF field in a controlled way. These sources produce very high RF powers up to several megawatts or substantial power at very high frequencies, up to several terahertz.

6.2 Mixer

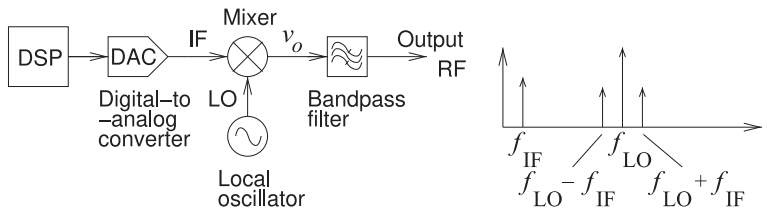
Frequency conversion, mixing or heterodyning, is the process of converting information at one frequency (present in the form of a modulated carrier) to another frequency. The second frequency is either higher, in the case of frequency **up-conversion**, where it is more easily transmitted, or lower, when mixing is called frequency **down-conversion**, where it is more easily captured. The mixer types are shown in Figure 6-1. Capture of the down-converted signal is nearly always by an ADC. Frequency conversion can occur with any nonlinear element.

Consider the information signal flow in the down-converter in Figure 6-1(a). (A similar discussion applies to the up-converter in Figure 6-1(b).) From the left, a modulated RF signal centered at f_{RF} is presented to a mixer that is pumped by a large LO signal at f_{LO}. The intended function of the mixer is to convert the information on the modulated RF to a lower intermediate frequency (IF) centered at $f_{\mathrm{IF}} = |f_{\mathrm{RF}} - f_{\mathrm{LO}}|$. The spectrum of the mixer, shown on the right on Figure 6-1(a), has another tone, f_{IM} called the image tone. The image at f_{IM} is an interferer, as it is also down-converted to the IF since $f_{\mathrm{IF}} = |f_{\mathrm{IM}} - f_{\mathrm{LO}}|$. Noise at the image is also down-converted to the IF. For the up-converter the high-power noise coming from the power amplifier at the image would be transmitted. Mixer design therefore must consider how image and noise are handled, as well as the efficiency of the conversion process.

In Figure 6-2(a) a nonlinear device is driven by two signals at ω_M and ω_C. The larger signal, the LO, is also called the **pump** and the other signal is called the RF. The spectrum of the signals present in the circuit is shown in Figure 6-2(c). In this mixer the aim is to produce a signal at the difference frequency (or IF) with the same modulation, and hence the same information, as the original RF signal. The transistor mixer shown in Figure 6-3 uses filtering to separate the RF, LO, and IF components.



(a) Down-conversion



(b) Up-conversion

Figure 6-1: Frequency conversion using a mixer.

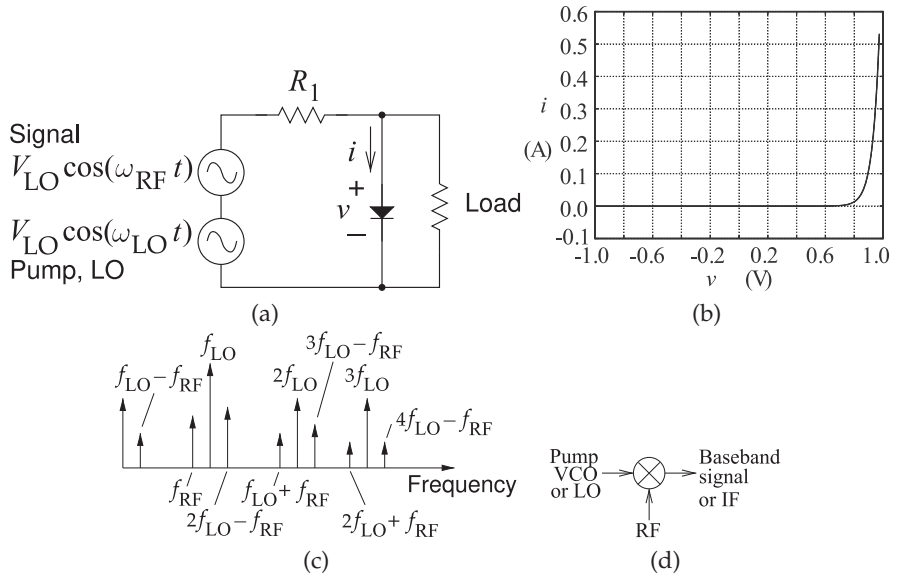


Figure 6-2: Diode mixer: (a) circuit; (b) diode current-voltage characteristic; (c) spectrum across the nonlinear device; and (d) schematic symbol for a mixer.

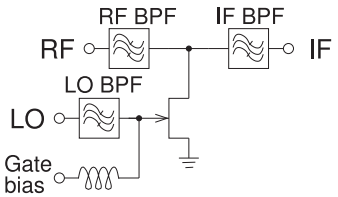


Figure 6-3: Single-ended FET mixer with LO, RF, and IF bandpass filters.

6.2.1 Mixer Analysis

A mixer can be designed around any nonlinear device [1]. Using an operational amplifier, an ideal multiplier can be designed that will multiply two signals, each described as cosinusoids. If the LO cosinusoid is $\cos(\omega_1 t)$ and the input RF cosinusoid is $\cos(\omega_2 t)$, the multiplication of the two signals will produce an output (using the trigonometric identity $\cos(A)\cos(B) = \frac{1}{2}[\cos(A-B) + \cos(A+B)]$):

$$\begin{aligned} y(t) &= [\cos(\omega_1 t)] \cdot [\cos(\omega_2 t)] \\ &= \tfrac{1}{2}\left\{\cos[(\omega_1 - \omega_2)t] + \cos[(\omega_1 + \omega_2)t]\right\}. \end{aligned} \tag{6.1}$$

This has two components: one at the radian frequency $(\omega_1 - \omega_2)$ and the other at $(\omega_1 + \omega_2)$. If the LO and RF signal are close, then the component at $(\omega_1 - \omega_2)$ will be at a frequency much lower than either the LO or the RF, and the component at $(\omega_1 + \omega_2)$ will be at almost twice the input frequencies. Appropriate filtering will choose one of these components depending on whether the application is up-conversion or down-conversion. At microwave frequencies, circuits that do not realize ideal multiplication must be used. This section considers what happens when two components are applied to an arbitrary nonlinearity described by a low-order polynomial. The result is that a large number of tones will be generated in the mixing process. Balanced circuit designs can significantly reduce many of these tones, greatly reducing the filtering required to select a particular output.

A two-tone input

$$x(t) = |X_1|\cos(\omega_1 t + \phi_1) + |X_2|\cos(\omega_2 t + \phi_2)$$

can be written using complex notation as

$$x(t) = \tfrac{1}{2}\left[X_1 e^{\jmath\omega_1 t} + X_1^* e^{-\jmath\omega_1 t} + X_2 e^{\jmath\omega_2 t} + X_2^* e^{-\jmath\omega_2 t}\right].$$

Note that the coefficient of the positive exponential frequency component is one-half that of the phasor. Thus the phasor of the ω_1 component is $X_1 = |X_1|e^{(\jmath\phi_1)}$ and the phasor of the ω_2 component is $X_2 = |X_2|e^{(\jmath\phi_2)}$. The first three powers of x can be easily expanded manually; for example, expanding x^2 gives

$$\begin{aligned} x^2(t) = \left(\tfrac{1}{2}\right)^2 \Big[& X_1^2 e^{\jmath 2\omega_1 t} + 2X_1X_1^* + 2X_1X_2 e^{\jmath(\omega_1+\omega_2)t} + 2X_1X_2^* e^{\jmath(\omega_1-\omega_2)t} \\ & + (X_1^*)^2 e^{-\jmath 2\omega_1 t} + 2X_1^*X_2 e^{\jmath(\omega_2-\omega_1)t} + 2X_1^*X_2^* e^{-\jmath(\omega_1+\omega_2)t} + X_2^2 e^{\jmath 2\omega_2 t} \\ & + 2X_2X_2^* + (X_2^*)^2 e^{-\jmath 2\omega_2 t}\Big], \end{aligned} \tag{6.2}$$

and similarly, expanding x^3 yields

$$\begin{aligned} x^3(t) = \left(\tfrac{1}{2}\right)^3 \Big[& X_1^3 e^{\jmath 3\omega_1 t} + 3X_1^2X_1^* e^{\jmath\omega_1 t} + 3X_1^2X_2 e^{\jmath(2\omega_1+\omega_2)t} \\ & + 3X_1^2X_2^* e^{\jmath(2\omega_1-\omega_2)t} + 3X_1(X_1^*)^2 e^{-\jmath\omega_1 t} + 6X_1X_1^*X_2 e^{\jmath\omega_2 t} \\ & + 6X_1X_1^*X_2^* e^{-\jmath\omega_2 t} + 3X_1X_2^2 e^{\jmath(\omega_1+2\omega_2)t} \\ & + 6X_1X_2X_2^* e^{\jmath\omega_1 t} + 3X_1(X_2^*)^2 e^{\jmath(\omega_1-2\omega_2)t} + (X_1^*)^3 e^{-\jmath 3\omega_1 t} \\ & + 3(X_1^*)^2X_2 e^{\jmath(\omega_2-2\omega_1)t} + 3(X_1^*)^2X_2^* e^{-\jmath(2\omega_1+\omega_2)t} + 3X_1^*X_2^2 e^{\jmath(2\omega_2-\omega_1)t} \\ & + 3X_1^*(X_2^*)^2 e^{-\jmath(\omega_1+2\omega_2)t} + X_2^3 e^{\jmath 3\omega_2 t} + 6X_1^*X_2^*X_2 e^{-\jmath\omega_1 t} \\ & + 3X_2^2X_2^* e^{\jmath\omega_2 t} + 3X_2(X_2^*)^2 e^{-\jmath\omega_2 t} + (X_2^*)^3 e^{-\jmath 3\omega_2 t}\Big], \end{aligned} \tag{6.3}$$

so that the output of the cubic equation,

$$y(t) = a_0 + a_1x(t) + a_2x^2(t) + a_3x^3(t),$$

can be calculated for a two-tone input. Table 6-1 lists these phasors and groups them by frequency. A more general approach is described in [2].

The phasors of the various intermodulation products resulting from x, x^2, and x^3 can be taken as the coefficients of the positive exponential frequency components after the factor of two correction required for terms other than DC [2]. Terms of the same frequency are summed to obtain the output at a particular frequency. For example, the phasor output at $(2\omega_1 - \omega_2)$ is given by the sum of three intermodulation products:

$$Y_{(2\omega_1-\omega_2)} = a_3\left(\tfrac{3}{4}\right)X_1^2X_2^*. \tag{6.4}$$

So the level of the output of a mixer, here $Y_{(2\omega_1-\omega_2)}$ at radian frequency $(2\omega_1-\omega_2)$, is related directly to the strength of the LO signal (with amplitude $|X_1|$), the strength of the nonlinearity (captured by the a_n coefficients), and the level of the input signal $|X_2|$. Unfortunately, low-order power series analysis as used here is not sufficient to model practical mixers, and computer-aided modeling tools are necessary. However, the manual analysis enables operation to be understood and architectures to be developed that intrinsically have the desired characteristics.

Intermodulation product	Frequency	Order
$\frac{1}{2}X_1X_1^*$	0	2
$\frac{1}{2}X_2X_2^*$	0	2
$2\frac{1}{2}X_1 = X_1$	ω_1	1
$2(\frac{1}{2})^3 3X_1^2X_1^* = \frac{3}{4}X_1^2X_1^*$	ω_1	3
$2(\frac{1}{2})^3 6X_1X_2X_2^* = \frac{3}{2}X_1X_2X_2^*$	ω_1	3
$2\frac{1}{2}X_2 = X_2$	ω_2	1
$2(\frac{1}{2})^3 3X_2^2X_2^* = \frac{3}{4}X_2^2X_2^*$	ω_2	3
$2(\frac{1}{2})^3 6X_1X_1^*X_2 = \frac{3}{2}X_1X_1^*X_2$	ω_2	3
$2(\frac{1}{2})^2X_1^2 = \frac{1}{2}2X_1^2$	$2\omega_1$	2
$2(\frac{1}{2})^2X_2^2 = \frac{1}{2}X_2^2$	$2\omega_2$	2
$2(\frac{1}{2})^3X_1^3 = \frac{1}{4}X_1^3$	$3\omega_1$	3
$2(\frac{1}{2})^3X_2^3 = \frac{1}{4}X_2^3$	$3\omega_2$	3
$2\frac{1}{2}X_1X_2 = X_1X_2$	$\omega_1 + \omega_2$	2
$2\frac{1}{2}X_1X_2^* = X_1X_2^*$	$\omega_1 - \omega_2$	2
$2(\frac{1}{2})^3 3X_1^2X_2 = \frac{3}{4}X_1^2X_2$	$2\omega_1 + \omega_2$	3
$2(\frac{1}{2})^3 3X_1^2X_2^* = \frac{3}{4}X_1^2X_2^*$	$2\omega_1 - \omega_2$	3
$2(\frac{1}{2})^3 3X_1X_2^2 = \frac{3}{4}X_1X_2^2$	$\omega_1 + 2\omega_2$	3
$2(\frac{1}{2})^3 3X_1^*X_2^2 = \frac{3}{4}X_1^*X_2^2$	$2\omega_2 - \omega_1$	3

Table 6-1: The intermodulation products resulting from x, x^2, and x^3, where x is a two-tone signal, showing only the positive frequencies. The first column gives the complex amplitudes (phasors) of the frequency components. (The order is the power of x.)

6.2.2 Mixer Performance Parameters

The main characteristics that define the performance of a mixer are the conversion gain or loss, and the noise figure [3, 4]. Mixing results from a nonlinear process that generates many tones and not just the ones of interest. Consequently, additional parameters are used to describe the performance of a mixer, and these derive from the generation of the additional tones. The mixer performance parameters are as follows:

Conversion loss: This is the ratio of the available power of the input signal to that of the output signal after mixing. It is usually expressed in decibels. In the diode mixer shown in Figure 6-5, the conversion loss is

$$L_C = \frac{P_{\text{in}}(\text{RF})}{P_{\text{out}}(\text{IF})}. \tag{6.5}$$

In decibels, the conversion loss is

$$L_C|_{\text{dB}} = 10\log_{10}\left[\frac{P_{\text{in}}(\text{RF})}{P_{\text{out}}(\text{IF})}\right]. \tag{6.6}$$

Noise figure (NF): The NF is 10 times the log of the noise factor F. The noise factor is the ratio of the SNR at the RF input to the SNR at the IF output (using the input noise generated by a resistor at standard temperature, 290 K).

However, there are qualifications for mixers. The first is that two inputs (at the RF and image frequencies) can produce noise and signal power at the IF. The double-sideband (DSB) NF includes signal and noise contributions from both the RF and the image frequencies. This is the situation with radiometry and astronomy, where the signal is at both sidebands. Single-sideband (SSB) NF includes the input signal at the RF only, but includes noise originating at both the RF and at the image frequencies. This is the situation for mixers in communications where the signal is only at one sideband.

Mixers can have substantial excess noise, as noise that is offset in frequency from the LO and its harmonics by the magnitude of the IF frequency will be down-converted to the IF for a down-converter, or converted to the RF in an up-converter. This process is sometimes called **noise folding**.

In summary, two noise figures are used with mixers, SSB NF and DSB NF. Which to use depends on the system in which the mixer is embedded.

Image rejection: If a mixer has an LO of 1.1 GHz and an RF of 1.5 GHz, then the IF will be at 400 MHz. This IF can also be generated by the image signal at 700 MHz. Numerically the image is the reflection of the RF in the LO. Image rejection refers to the ability of a mixer to reject the image signal. This can be achieved, for example, by using an input RF bandpass filter.

If the applied image and intended signal powers are the same and the level of the output signal (at the IF) produced by the intended RF signal is P_{out}, and that produced by the image signal is $P_{\text{out,image}}$, then the **image rejection ratio** (**IRR**) is

$$\text{IRR} = \frac{P_{\text{out}}}{P_{\text{out,image}}}. \tag{6.7}$$

This is typically expressed in decibels and

$$\text{IRR}|_{\text{dB}} = 10\log_{10}\left(\frac{P_{\text{out}}}{P_{\text{out,image}}}\right). \tag{6.8}$$

EXAMPLE 6.1 Mixer Calculations

A mixer has an LO of 10 GHz. The mixer is used to convert a signal at 10.1 GHz to an IF at 100 MHz, and has a conversion loss, L_c of 3 dB and an image rejection of 20 dB. Two signals are presented to the mixer, one at 10.1 GHz with a power of 100 nW and the other at 9.9 GHz with a power of 1 μW.

(a) What is the output power of the (intended) signal at the IF?
(b) What is the signal-to-interference ratio at the IF (ignoring noise)?

Solution:

(a) $L_c = 3$ dB = 2 and from Equation (6.5) the output power at IF of the intended signal is

$$P_{\text{out}} = P_{\text{in}}(\text{RF})/L_c = 100\text{ nW}/2 = 50\text{ nW} = -43\text{ dBm}. \tag{6.9}$$

(b) Interference at the IF comes from the down-converted image signal. IRR = 20 dB = 100. If the applied powers of the intended signal and the image signal are the same, from Equation (6.7),

$$\text{IRR} = \frac{P_{\text{out}}}{P_{\text{out,image}}} \quad \text{i.e.} \quad P_{\text{out,image}} = \frac{P_{\text{out}}}{\text{IRR}} \tag{6.10}$$

This must be modified to account for the difference in the applied power levels and

$$P_{\text{out,image}} = \frac{P_{\text{out}}}{\text{IRR}} \frac{P_{\text{in}}(\text{RF, image})}{P_{\text{in}}(\text{RF})} = \frac{(50\ \text{nW}) \cdot (1\ \mu\text{W})}{100 \cdot (100\ \text{nW})} = \frac{(50\ \text{nW}) \cdot (1000\ \text{nW})}{100 \cdot (100\ \text{nW})} = 5\ \text{nW}.$$

The signal-to-interference ratio is

$$\text{SIR} = \frac{P_{\text{out}}}{P_{\text{out,image}}} = \frac{50\ \text{nw}}{5\ \text{nw}} = 10 = 10\ \text{dB}. \tag{6.11}$$

6.2.3 Mixer Waveforms

This section presents the waveforms of diode mixers. Diode mixers, such as the balanced mixer in Figure 6-4, are bilateral. In a communication transceiver such a mixer can be used for both the receive and transmit functions, greatly simplifying filtering requirements. Most mixers in cell phones, however, are based on transistors, but the underlying architecture corresponds to a single-diode mixer or, more commonly, the diode ring double-balanced mixer.

The first mixer to be considered is the single-ended mixer shown in Figure 6-5(a). Filters are not included thus revealing the full complexity of the spectra. The transient waveform across the diode (measured at the test point) is shown in Figure 6-5(b). The turn-on transient is due to both the turn-on ramp of the LO and RF sources, and to the internal capacitance of the diode. The spectra of the signal at the test point is shown in Figure 6-5(c and d). The desired signal here is the IF, which is at the difference frequency of the LO and RF (i.e., at 400 MHz). Extracting just the IF in the response shown in Figure 6-5(d) would require significant filtering. The level of the IF tone here is 11.5 mV_{peak}. If a lossless 400 MHz filter is attached to the test point and the IF load (on the other side of the filter) is 50 Ω, the IF power delivered to the load is $P_{\text{out}} = \frac{1}{2}(11.5\ \text{mV})^2/(50\ \Omega) = 1.322\ \mu\text{W}$. The available RF power is $P_{\text{in}} = \frac{1}{2}(100\ \text{mV})^2/(50\ \Omega) = 100\ \mu\text{W}$. So the conversion loss is

$$L_C = \frac{100\ \mu\text{W}}{1.332\ \mu\text{W}} = 75.08 = 18.8\ \text{dB}. \tag{6.12}$$

Now consider the diode ring mixer shown in Figure 6-6. Waveforms and spectra at the LO, IF, and RF test points are shown in Figure 6-7. Note how relatively clean the spectra are at each of the test points. The reduction in clutter is the result of the balanced circuit. The circuit in Figure 6-6 is known as a double-balanced mixer, and the same concept can be used with transistor mixers.

Focusing on just the spectra at the IF test point (Figure 6-7(e)), it can be seen that the LO signal and its harmonics are nonexistent. The components other than the 400 MHz IF are other sum and difference products of the RF and LO. However, compared to the spectra of the single-ended diode mixer

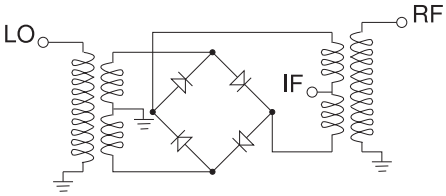
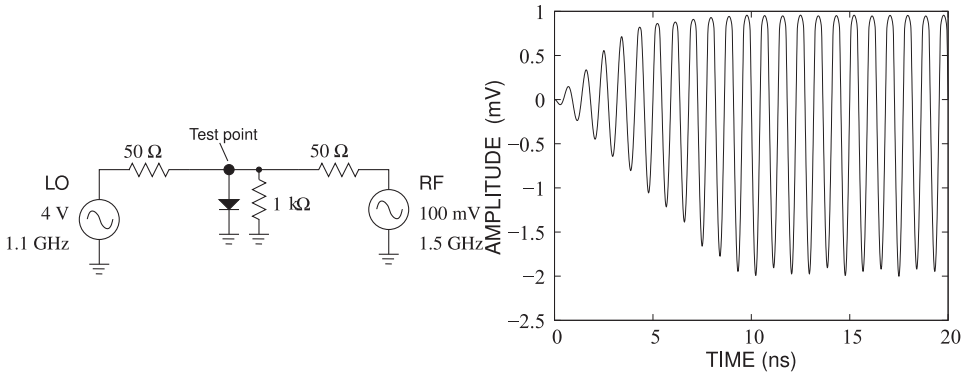
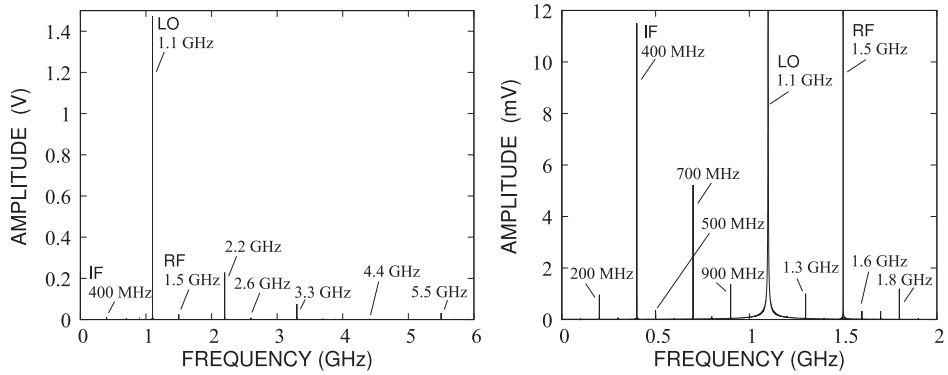


Figure 6-4: Diode ring double-balanced mixer.



(a) Schematic (b) Waveform at the test point



(c) Spectrum at the test point (d) Expanded spectrum at the test point

Figure 6-5: Single-ended diode mixer.

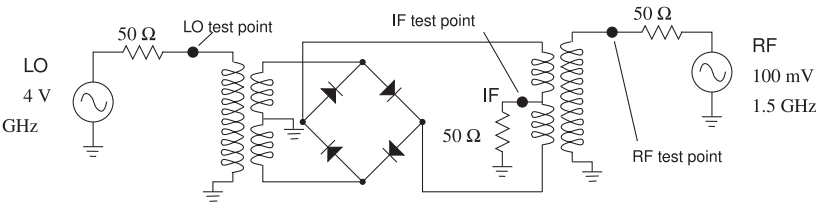
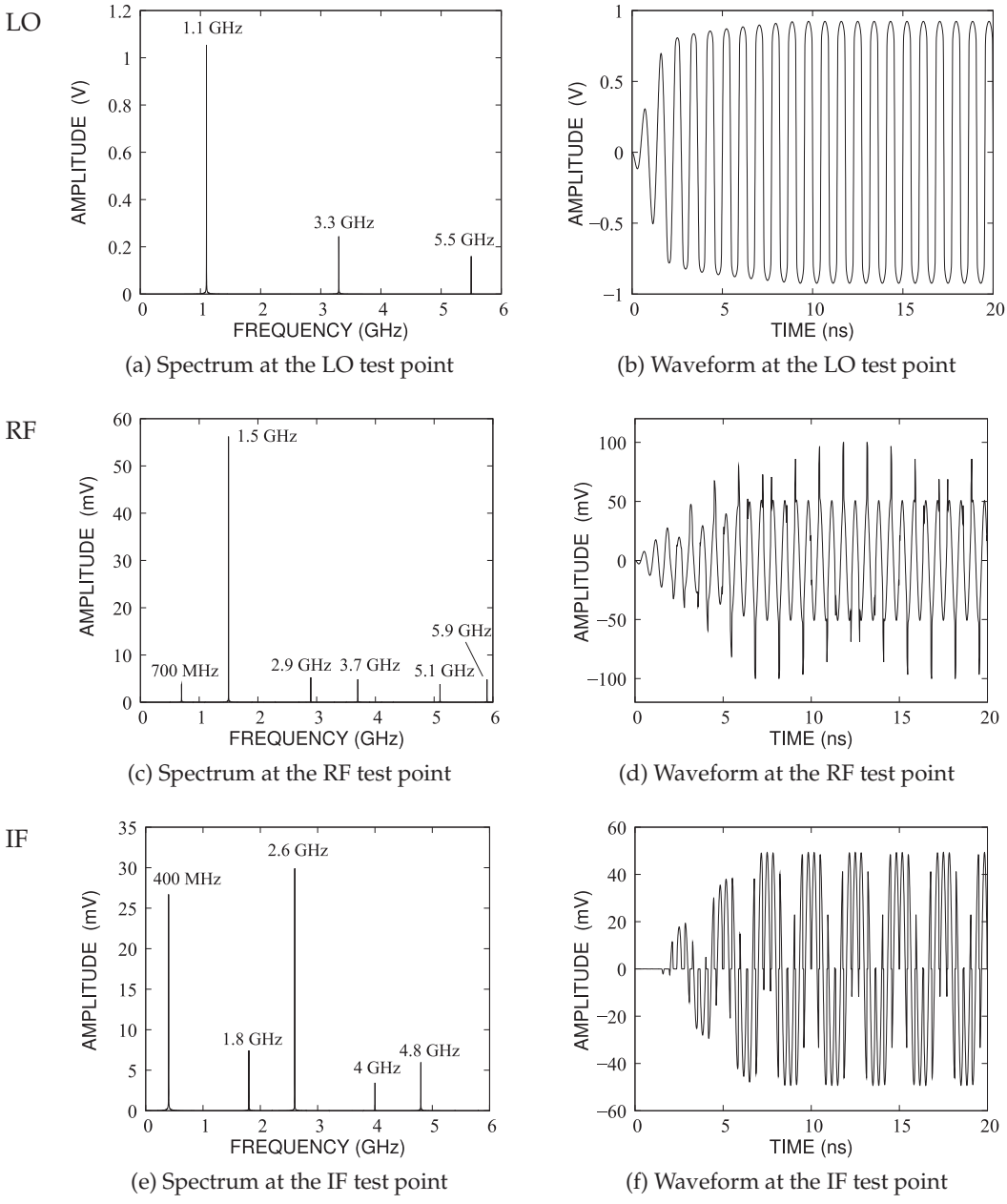


Figure 6-6: Diode double-balanced mixer schematic.

in Figure 6-5(c and d), many of these are eliminated as well. This is a very attractive circuit, as it significantly reduces the specifications required for an output filter. This is one of the special characteristics of RF and microwave design. Much can be gained by being creative—a designer gets better with experience. The level of the IF signal at the RF test point is 26.7 $\mathrm{mV_{peak}}$. The IF power delivered to the 50 Ω load is $P_{\mathrm{out}} = \frac{1}{2}(26.7\ \mathrm{mV})^2/(50\ \Omega) = 7.126\ \mu\mathrm{W}$. The available RF power is $P_{\mathrm{in}} = \frac{1}{2}(100\ \mathrm{mV})^2/(50\ \Omega) = 100\ \mu\mathrm{W}$. So the conversion loss is

$$L_C = \frac{100\ \mu\mathrm{W}}{7.126\ \mu\mathrm{W}} = 14.03 = 11.47\ \mathrm{dB}. \tag{6.13}$$

The conversion loss is much lower than was obtained with the single-ended mixer. In part this is because power was not dissipated in a large number of spurious tones.



(a) Spectrum at the LO test point

(b) Waveform at the LO test point

(c) Spectrum at the RF test point

(d) Waveform at the RF test point

(e) Spectrum at the IF test point

(f) Waveform at the IF test point

Figure 6-7: Waveforms and spectra of the double-balanced diode ring mixer of Figure 6-6.

6.2.4 Switching Mixer

The analysis of a mixer in Section 6.2.1 modeled the mixing element using a polynomial. With a large LO, an alternative view of a mixer is to consider it as a switching device in which the conductance of the mixing element is periodically switched by the LO.

A simple switching diode mixer is shown in Figure 6-8(a). If the LO level is large, then the mixer can be modeled by the equivalent circuit shown in Figure 6-8(b) with a periodically varying conductance, $g_{\rm LO}(t)$. This is a square wave with the same frequency and period as the applied LO sinewave. The diode effectively looks like a switch with the conductance alternating between a maximum value g_m and zero (see Figure 6-8(c)). The first few terms of the Fourier series expansion of $g_{\rm LO}(t)$ are

$$g_{\rm LO}(t) = g_m \left[\frac{1}{2} + \frac{1}{\pi}\cos(\omega_{\rm LO}t) + \frac{2}{3\pi}\cos(3\omega_{\rm LO}t) + \ldots\right]. \tag{6.14}$$

The small RF signal interacts with the switching conductance so that the output current is

$$i(t) = g_{\rm LO}(t)\left[v_{\rm RF} - i_x(t)R\right]. \tag{6.15}$$

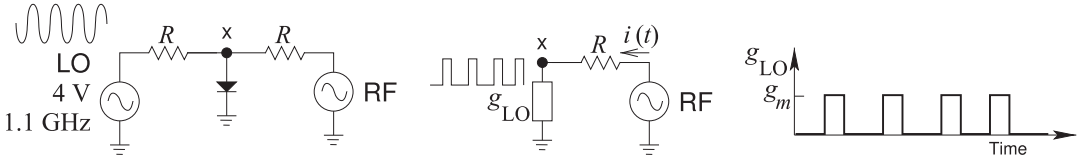
Ignoring R, then the IF current, where $\omega_{\rm IF} = \omega_{\rm RF} - \omega_{\rm LO}$, at the point x is

$$\begin{aligned} i_{\rm IF}(t) &= g_{\rm LO}(t)v_{\rm RF}(t) = \frac{g_m}{\pi}\cos(\omega_{\rm LO}t)v_{\rm RF}\cos(\omega_{\rm RF}t) \\ &= \frac{g_m v_{\rm RF}}{2\pi}\left\{\cos[(\omega_{\rm LO} - \omega_{\rm RF})t] + \cos[(\omega_{\rm LO} + \omega_{\rm RF})t]\right\}. \end{aligned} \tag{6.16}$$

Typically the IF signal at $(\omega_{\rm LO} - \omega_{\rm RF})$ would be extracted through a bandpass or lowpass filter that allows only the IF component to pass and an IF voltage is realized as the current passes through a load resistor.

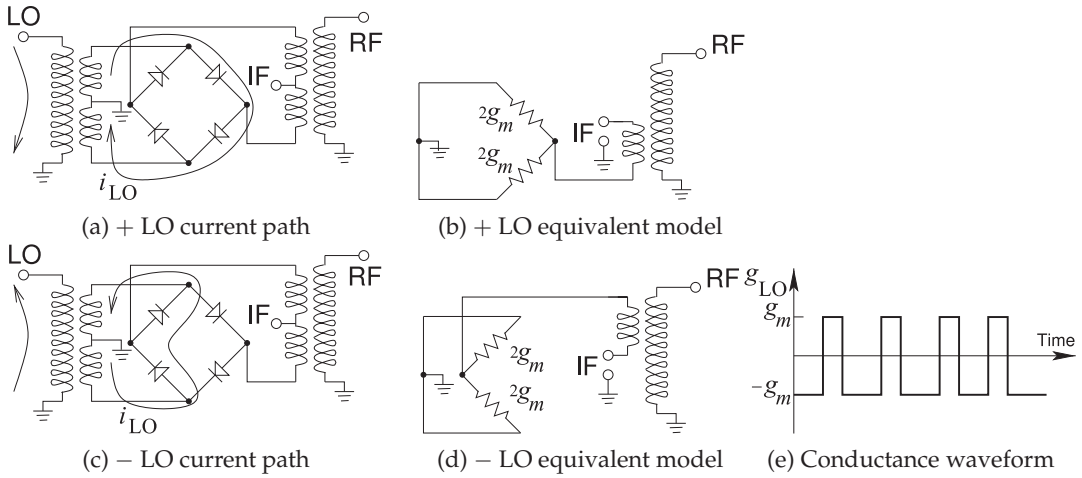
One of the advantages of the switching mixer is that performance is relatively insensitive to the level of the LO. The LO could be a sinewave and still the variation of the conductance would be close to being a square wave. So then the design of the mixer specifically is to develop a square variation of the conductance.

Switching mixers can be realized using other circuits. One of these is the diode ring mixer shown in Figure 6-9. Here the center-tapped transformers produce differential LO and RF signals. The large LO is transformed by the tapped transformer to produce a large differential signal that turns pairs of diodes on in sequence. During the positive half of the LO cycle, the



(a) Diode mixer (b) Equivalent model (c) Conductance waveform

Figure 6-8: Switching diode mixer; $g_{\rm LO}$ is a conductance.



(a) + LO current path (b) + LO equivalent model

(c) − LO current path (d) − LO equivalent model (e) Conductance waveform

Figure 6-9: Diode ring mixer as a switching mixer. g_m is a conductance.

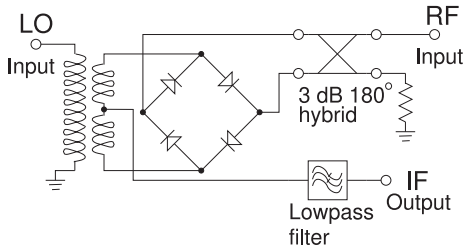


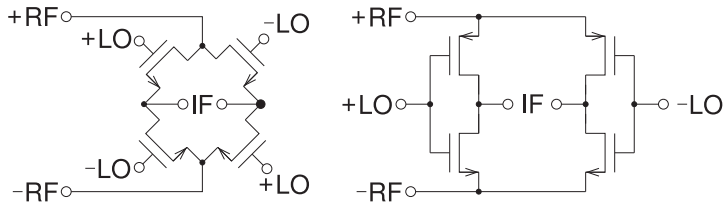
Figure 6-10: Diode ring mixer implemented using a hybrid to produce a pseudo-balanced signal.

two right-hand diodes in Figure 6-9(a) pass current limited by the source impedance of the LO source and each diode has a conductance $2g_m$, as seen in Figure 6-9(b). During the negative half-cycle of the LO, the left-hand diodes pass current and each has conductance $2g_m$, see Figure 6-9(c and d). When not forward biased, the diodes appear as open circuits. This process results in the conductance waveform shown in Figure 6-9(e). An important characteristic of this waveform is that it is symmetrical even when nonidealities are considered.

An RF implementation of the diode mixer is shown in Figure 6-10. The 180° hybrid replaces the transformer to distribute oppositely phased RF signals to the diode ring. The IF is now taken from the center tap of the LO transformer and is passed through a lowpass filter to remove all signals other than the IF. This is a down-converting mixer. If the mixer is an up-converting mixer, then the lowpass filter would be replaced by a bandpass filter.

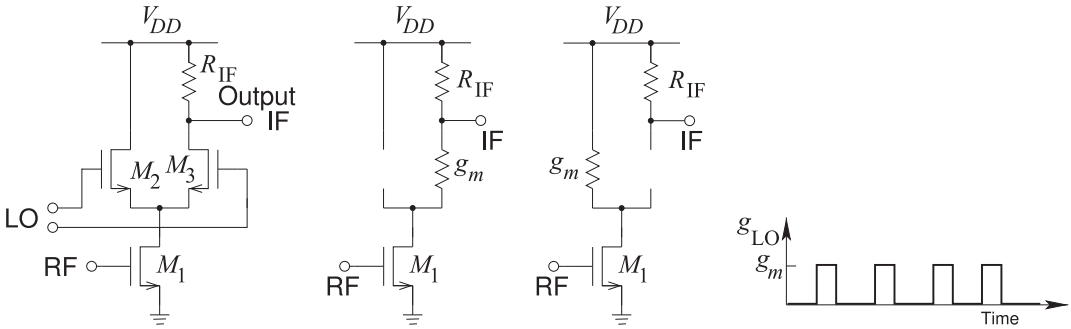
Transistor-based mixer implementations using the ring mixer concept are shown in Figure 6-11. These circuits are used in monolithic ICs with the differential signals available from preceding stages and the IF output is also differential. The transistors operate as switches that are controlled by the LO signal and the conductance waveform is as for the diode mixer (i.e., as in Figure 6-9(e)). Another switching mixer is the transistor **commutating mixer** shown in Figure 6-12.

Traditionally the the biggest issue with switching mixers was is the substantial LO power required and the limited dynamic range of the mixer in



(a) MOS ring mixer (b) CMOS H-bridge ring mixer

Figure 6-11: Transistor ring mixers.



(a) Commutating mixer (b) + LO model (c) − LO model (d) Conductance

Figure 6-12: Commutating transistor mixer.

converting RF to IF (for a down-converter) or IF to RF (for an up-converter). This situation changes with the 32 nm and lower CMOS processing nodes. This has resulted in CMOS switching mixers being important elements of RFICs and these have superior dynamic range.

6.2.5 Subharmonic Mixer

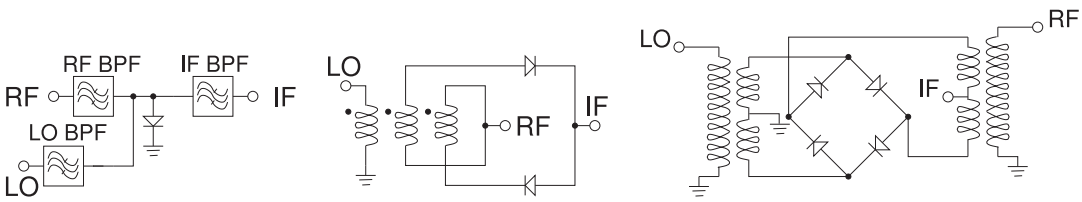
If the nonlinear element of a mixer is particularly strong, or perhaps the drive level is very high, mixing will occur when the frequency of the LO drive of the mixer is a subharmonic of the the effective LO frequency. Consider the switching mixers discussed in Section 6.2.4 and the switching conductance, Figure 6-12(c), of the commutating mixer in Figure 6-12(a). With this mixer the RF signal effectively mixes with an harmonic of the applied LO. Here these harmonics are odd harmonics and so a commutating mixer driven by an LO signal at 5 GHz will down-convert and RF signal at 15.1 MHz to an IF frequency of 100 MHz. One of the advantages of a subharmoinic mixer is that there is not a large 15 GHz signal which could be a problem as it could leak into the front of a receiver and possibly radiate unintentionally. Another advantage is that with a conventional down-conversion mixing approach where the LO and RF are close in frequency, an LO at 15 GHz, in previous example, would need to be generated from the 5 GHz signal. This would consume considerable power. So it could be more power efficient to use a subharmonic mixer rather than a frequency tripler and a conventional mixer. A subharmonic mixer could be the preferred design option especially for down-converting millimeter-wave signals. At high millimeter-wave frequencies (e.g. > 200 GHz) the mixer is often a passive nonlinear element.

6.3 Single-Ended, Balanced, and Double Balanced Mixers

Single-ended, balanced, and double-balanced diode mixers are shown in Figure 6-13. The simplest type of mixer that has been considered is the single-ended or **unbalanced diode mixer** of Figure 6-2(a). This is shown as a down-converting mixer where the RF signal at f_{RF} mixes with the LO at f_{LO} to produce a down-converted IF signal at $f_{\mathrm{LO}} - f_{\mathrm{RF}}$. The image signal at $f_{\mathrm{LO}} + (f_{\mathrm{LO}} - f_{\mathrm{RF}}) = 2f_{\mathrm{LO}} - f_{\mathrm{RF}}$ will be down-converted to the same IF signal. It is of course not desirable to down-convert the image signal and it could be prevented by using filtering as shown, but the use of filters makes the filter quite large. There are many spurious frequencies with the single-ended mixer as seen in Figure 6-5. Most of these can be removed with filtering but using filters is undesirable and certainly not compatible with on-chip implementations. The solution is to use symmetry to cancel out many of the spurious tones. Figure 6-13(b and c) shows two types of symmetry, balanced and double balanced mixers. Another level of symmetry could be used (applying it to the IF) in the double balanced mixer and then it could be called a **triple balanced mixer**. The triple balanced mixers is nearly always referred to as a double-balanced mixer. One cannot say whether a mixer is balanced, or even double balanced without tracing through the cancellation process. With each increasing degree of symmetry greater linearity of the conversion process is obtained. The double balanced mixer in FET form is particularly attractive as the FETs do not need to be biased. Except at very high millimeter-wave frequencies and terahertz frequencies, all mixers today are of the double or triple balanced form. However there can be unintentional mixing so the unbalanced mixer should be studied so that unwanted mixing can be identified.

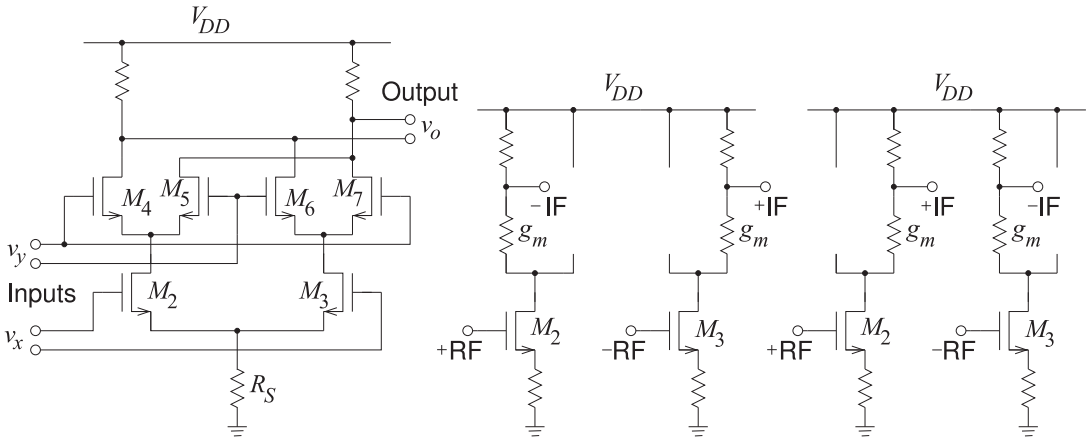
6.3.1 RFIC Mixers

RFIC mixers below millimeter-wave frequencies utilize the transistor as a switch either with finite conductance when the switch closed and virtually zero conductance when the switch is open. The first mixer to be considered is based on the **commutating mixer** shown in Figure 6-12.



(a) Unbalanced (b) Balanced (c) Double balanced

Figure 6-13: Unbalanced (also known as single-ended), balanced, and double-balanced down-conversion diode mixersthat mix an LO with an RF to produce a lower frequency IF. bandpass filters.



(a) Gilbert cell (b) + LO model (c) − LO model

Figure 6-14: Model of the Gilbert cell and equivalent circuit models. The circuit is used as a mixer where v_y is the LO and v_x is the RF. (This is referred to as a double balanced mixer but strictly it is a triple balanced mixer as there are RF, IF, and LO symmetries resulting in a high level of cancellation of spurious tones.)

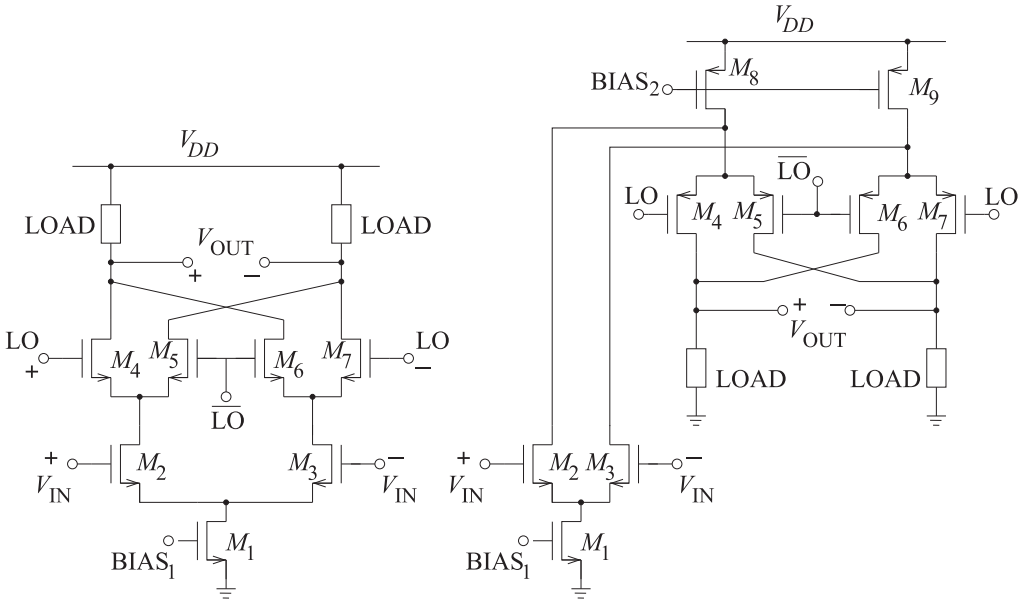
6.3.2 Gilbert Cell

The Gilbert cell builds on the commutating mixer circuit of Figure 6-12. The commutating mixer conducts current in the IF leg (with R_{IF} during only half of the LO waveform. An immediate improvement is obtained by modifying the circuit to provide switched conductance during both halves of the LO waveform. Also, the RF and IF in the commutating mixer are single-ended quantities whereas with a silicon IC differential signals are preferred. The simplest Gilbert cell is shown in Figure 6-14(a), which has all of the desired properties. M_2 and M_3 are designed as a differential amplifier pair with M_4 and M_5 functioning as time-varying loads for M_2, and M_6 and M_7 function as time-varying loads for M_3. This leads to the equivalent model of the Gilbert cell shown in Figures 6-14(b and c).

The Gilbert cell is not always used as a mixer. If both v_y and v_x are small signals, then the Gilbert cell becomes a good **analog multiplier** with the output $v_o = v_x \times v_y$. When used as a mixer, v_y is a large LO and the Gilbert cell is referred to as a Gilbert mixer. The Gilbert mixer is the functional core of virtually every integrated mixer, especially in CMOS technology where differential operation is preferred.

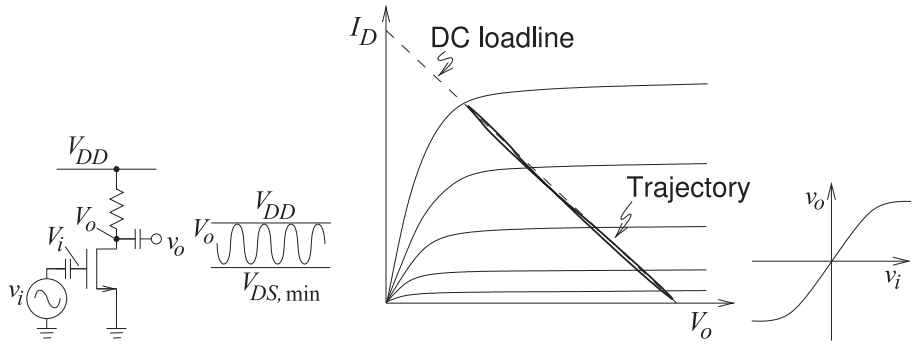
6.3.3 Integrated Gilbert Cell-Based Mixer

The performance of a Gilbert mixer is improved by replacing the source resistor, R_S, of M_2 and M_3 in the basic Gilbert cell of Figure 6-12 with a current source, as shown in Figure 6-15(a). This is the most common variation of the Gilbert mixer used in RFICs. A double-balanced Gilbert cell mixer is shown in Figure 6-15(a). When the LO is small, the Gilbert mixer has an almost ideal multiplier response with performance limited by how closely the transistors can be matched [5]. In RFICs the transistors can be



(a) Gilbert mixer (b) Folded cascode implementation

Figure 6-15: Gilbert cell double-balanced mixer.



(a) Class A amplifier (b) Loadline (c) Tanh-like characteristic

Figure 6-16: Simple transistor circuit (biasing not shown).

matched very well. When the LO is sufficiently large to produce a switching conductance, there is much greater tolerance to imperfect matching. This circuit has good performance and rejects RF and LO feed-through to the output, and the degree of rejection is dependent on the quality of the transistor match.

Before continuing the discussion the gross characteristics of a transistor circuit will be introduced. The simplest transistor circuit is the amplifier shown in Figure 6-16(a). With an RF input signal, the large capacitor at

the gate of the transistor becomes a short circuit and the trajectory of the transistor's voltages and currents follows a load line as shown in Figure 6-16(b). In the form of a transfer function, the characteristic of the amplifier is as shown in Figure 6-16(c). This type of characteristic is typical of transistor circuits with the saturation levels of the output voltage set by the supply voltage and ground. The characteristic shown in Figure 6-16(c) closely resembles the trigonometric tanh function and so the characteristic of a transistor circuit is often said to be tanh-like. A tanh-like response is also obtained with a Gilbert mixer circuit.

One problem with the Gilbert mixer circuit of Figure 6-15(a) is the reduced voltage swing resulting from three drain-source voltage drops between the supply rails. The classic technique for solving this problem is to use a folded cascode design. The folded Gilbert cell double-balanced mixer then becomes the circuit of Figure 6-15(b) [6, 7]. This merges a cascode amplifier design with the double-balanced Gilbert cell mixer in Figure 6-15(a). The result is a mixer that can have a larger voltage swing.

Gilbert mixers, as enhanced so far, can have poor linearity in the sense that the IF output is linear only for small RF input signals. This problem is exasperated by the trend to reduce the supply voltage of integrated circuits. The solution used is to replace each of the amplifying transistors (M_2 and M_3 in Figures 6-15(a and b)) by multiple transistors with progressively offset biasing that staggers the tanh-like transfer characteristic of a transistor circuit to realize an effective transistor with more linear characteristics [6, 8]. This design is used in the up-converting mixer shown in Figure 6-17, of a WCDMA transmitter RFIC. and the variation in biasing of the multiple amplifying transistors is achieved using sizing of each of the current source transistors in the bottom row. Alternatively, separate bias could be applied to each current source transistor. Also, the double-balanced Gilbert cell mixer in Figure 6-15(a) is duplicated to realize a quadrature mixer.

The CMOS mixer has evolved from the core concept based on the diode ring mixer, first replacing the diodes by transistors, and then introducing transistor structures that cope with limited voltage supply and transistor nonidealities. This process is the essence of RF integrated circuit design.

6.3.4 Ring-Based Mixers

Diode ring mixers were discussed in Section 6.2.4 and circuit implementations were shown in Figures 6-9 and 6-10. Transistor-based mixer implementations using the ring mixer concept are shown in Figure 6-18.

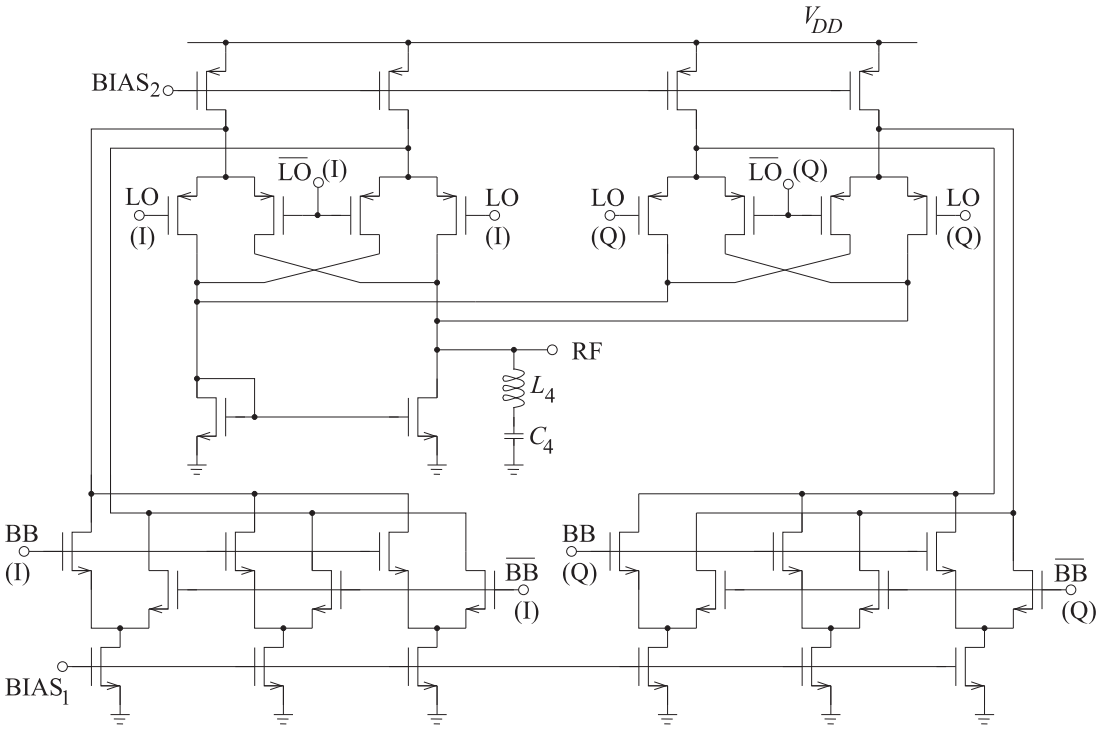


Figure 6-17: Quadrature up-converting mixer of a WCDMA transmitter. BB is the low-frequency baseband signal with (I) and (Q) components. $\overline{\text{BB}}$ is the inverse (negative) of BB. LO and $\overline{\text{LO}}$ are the differential local oscillator signal, and RF is the single-ended output signal. The LO is at 2 GHz, the baseband input frequency is 5 ± 1.72 MHz, and the output frequency is 1.95 MHz. After Yang and Gard [6] and Yang [7]. Copyright K. Gard and X. Yang, used with permission.

6.3.5 Summary

Mixer type	Feature	Disadvatages
Unbalanced mixer	Very simple. Suitable for high millimeter-wave and terahertz frequencies.	Requires bulky filters to separate signals. Large spurious signals.
Single balanced mixer	Suppression of many spurious tones. Suppression of amplitude-modulated noise on LO. Increased linearity compared to unbalanced mixer.	Requires high LO drive level
Double balanced mixer	Increased linearity compared to single-balanced mixer. All ports of the mixer are inherently isolated from each other. Can be implemented with passive mixing elements not requiring biasing. Inherently broadband.	Relies on symmetry and matched diodes and/or transistors. On RFIC calibration is used to balance mixer.
Triple balanced mixer	Provides increased linearity. The type most commonly used in RFICs although commonly referred to as a double balanced mixer.	Increasingly complex circuitry.

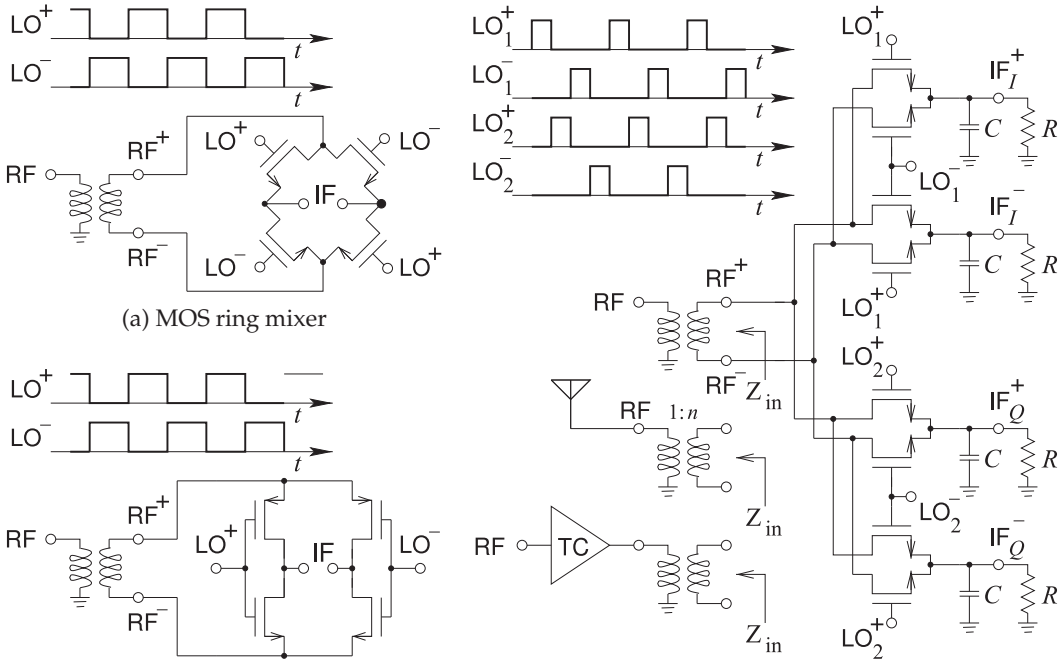


Figure 6-18: Transistor ring-type mixers.

6.4 Local Oscillator

In this section the local oscillator that is used to drive a mixer will be treated as a black box. The detailed design of oscillators is considered in Chapter 5 of [9].

6.4.1 Phase Noise in Local Oscillators

This section builds on the discussion of noise in Section 4.2. The performance of most RF and microwave systems is limited by oscillator noise. In an oscillator the noise close to the carrier, or oscillation frequency, is called flicker noise and it significantly affects system performance. (In RF amplifiers flicker noise is generated but it is of much less concern than it is with oscillators.) Noise close to the oscillation center frequency (tens of hertz to a few megahertz away for RF and microwave oscillators) manifests itself as random fluctuations of amplitude and phase of the carrier. The amplitude fluctuations are quenched by saturation in the oscillator and so are not of concern. Thus the important close-in noise is just phase noise. This phase noise increases the smaller frequency offset from the oscillation frequency.

The main use of an LO is to drive a mixer. Then phase noise on the LO is added to the phase of the signal converted by the mixer and so the LO phase noise becomes phase noise on the converted signal. In a digitally modulated communication system the components of the phase noise at the symbol rate and down to a few of its subharmonics have the most impact on the ability to demodulate signals. Phase noise outside the bandwidth of

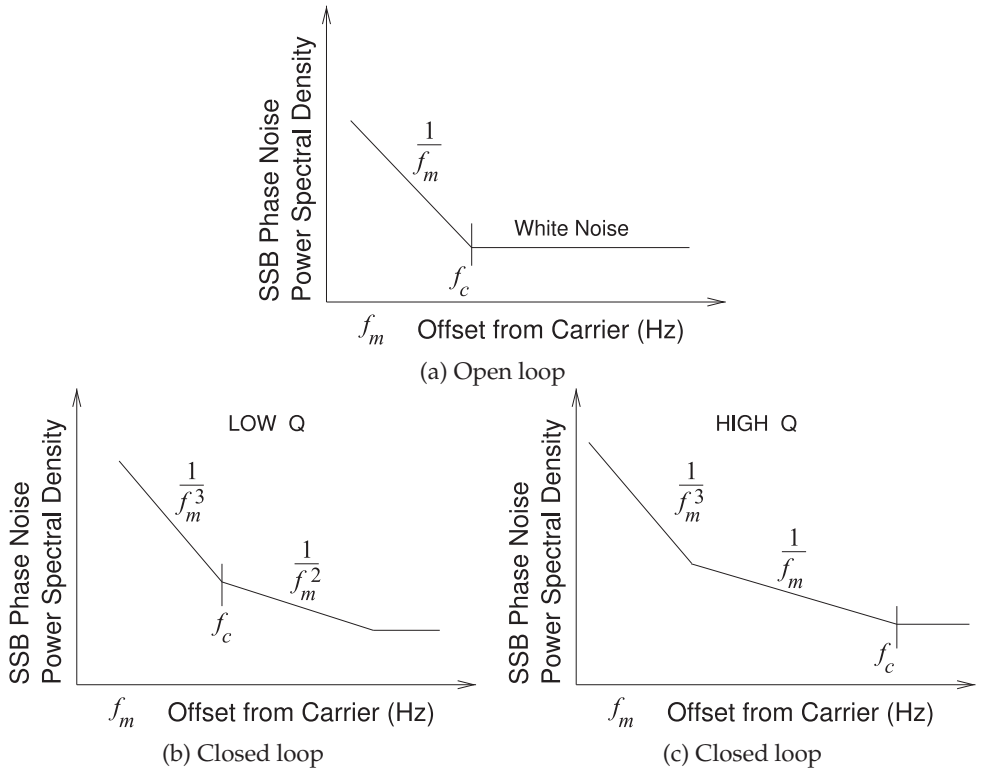


Figure 6-19: Log-log plot of oscillator noise spectra: (a) open-loop noise showing flicker noise ($1/f_m$) and white noise regions; (b) closed-loop noise with low-Q loop; and (c) closed-loop noise with high-Q loop.

the communication signal can be filtered out. Also phase noise slower than a few subharmonics below the symbol rate are compensated for in signal processing (e.g., using error correction codes).

Most RF oscillators comprise free-running oscillators whose oscillation frequency can be controlled by an applied DC or low-frequency voltage. Superior performance is obtained by comparing a scaled down version of an oscillator output to a high-precision reference oscillator such as a crystal oscillator. Without the feedback the oscillator is said to be an open-loop oscillator, and with feedback it is said to be a closed-loop oscillator.

For a free-running oscillator (i.e., an open-loop oscillator), the phase noise close to the **carrier** (i.e. the average oscillation frequency) is dominated by flicker phase noise, as shown in Figure 6-19(a). This describes the intrinsic noise property of the active device (and surrounding circuitry) and the white noise and flicker noise responses are clearly seen. When the loop is closed, the loop transfer characteristic shapes the noise response, producing noise that has regions close to the carrier that has a $1/f_m^3$ shape, and further from the carrier it varies as $1/f_m^2$ if the Q of the loop is low [10]. The switch from $1/f_m^3$ to $1/f_m^2$ dependence is at what is called the transition or crossover frequency, f_c (see Figure 6-19(b)). If the Q of the loop is high, the phase noise profile will transition from the $1/f_m^3$ regime directly to the

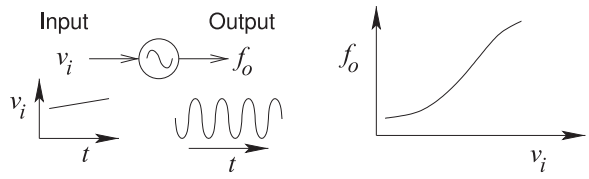


Figure 6-20: Voltage-controlled oscillator. (a) Schematic (b) Tuning curve

$1/f_m$ regime, and again the transition frequency is f_c (see Figure 6-19(c)). The transition frequency is usually around a few kilohertz to hundreds of kilohertz offset from the carrier for microwave transistors. A feedback analysis that describes how $1/f$ and white noise are converted to the $1/f_m^3$, $1/f_m^2$, or $1/f_m$ characteristics was developed by Leeson in 1966 [10] and others [11]. An advanced discussion of oscillator phase noise is given in Section 5.8 of [9] and the discussion here is limited but sufficient when using oscillators as modules. Here the traditional but approximate approach developed by Leeson is followed.

Phase noise was formally defined in Equation (4.60) and roughly it is the ratio of the phase noise power in a 1 Hz bandwidth of a single sideband (SSB) to the total signal power. This is measured at a frequency f_m offset from the carrier and denoted $\mathcal{L}(f_m)$ with the units of dBc/Hz (i.e., decibels relative to the carrier power per hertz).

The phase noise that is important in RF and microwave oscillators (having relatively low Q) is usually dominated by a $1/f_m^2$ shape. Then the phase noise at 1 MHz (a common frequency for comparing the phase noise performance of different oscillators) is related to the phase noise measured at f_m by

$$\mathcal{L}(1\text{ MHz}) = \mathcal{L}(f_m) - 10\log\left(\frac{1\text{ MHz}}{f_m}\right)^2. \tag{6.17}$$

Another commonly used quantitative assessment of oscillator performance is provided by the oscillator figure of merit, FOM_1, which accounts for the DC power consumed [12]:

$$\text{FOM}_1 = \mathcal{L}(f_m) - 10\log\left(\frac{1\text{ MHz}}{f_m}\right)^2 + 10\log\left(\frac{P_{\text{DC}}}{P_{\text{ref}}}\right), \tag{6.18}$$

where P_{ref} is conventionally taken as 1 mW and FOM_1 is referenced to the phase noise at 1 MHz.

6.5 Voltage-Controlled Oscillator (VCO)

As its name implies, a controlling voltage sets the frequency of the output of a VCO. The block diagram of a VCO is shown in Figure 6-20(a) where a slowly varying input signal, v_i, determines the frequency, f_0, of the signal produced by an oscillator. This is usually achieved by varying the capacitance of a varactor diode in a resonator.

One of the important characteristics of a VCO is the tuning curve, which plots the output frequency f_0 against the applied tuning voltage v_i as shown in Figure 6-20(b). Ideally the tuning curve is a straight line, but in practice it is shaped and the actual range over which the VCO is used is less than the full range supported. The tuning property is described by the **tuning constant**,

Table 6-2: Comparison of RF VCOs. Phase noise is worst case over tuning range; RF output power is the minimum. All oscillators are hybrids unless indicated by IC, denoting an integrated circuit. If f_m is not 1 MHz, then a $1/f^2$ dependence is assumed for the phase noise to calculate the phase noise at 1 MHz. The CMOS VCOs are quadrature VCOs producing two outputs $90°$ apart. After [13] with corrected FOM_1. ($P_{\text{ref}} = 1$ mW, $f_{\text{ref}} = 1$ MHz.)

f_0 GHz	f_{BW} MHz	P_{RF} dBm	P_{DC} mW	f_m MHz	$\mathcal{L}(f_m)$ dBm /Hz	$\mathcal{L}$(1 MHz) dBm /Hz	FOM_1 dBm /Hz	FOM_2 dBm /Hz	Reference
4.92	770	0	150	1	−128		−106	−157	SiGe HBT hybrid [13]
5.05	500	0	150	1	−130		−106	−155	SiGe HBT hybrid [13]
5.16	229	−0.43	24	1	−111		−98	−136	InGaP/GaAs HBT [14]
11.5	550	9		0.1	−91	−111		−138	GaAs MESFET [15]
9.33	440	3.3	30.5	1	−102		−87	−128	GaN HEMT [16]
6.40	150	5.5	173	0.1	−105	−125	−85	−127	SiGe HBT [17]
5.94	166	−4.0	8.1	1	−110		−94	−134	CMOS IC [18]
4.87	70	−4.0	4.8	1	−131		−124	−149	GaInP/GaAs HBT [19]
5.38	120	−4.0	12.8	1	−127		−108	−148	GaInP/GaAs HBT [20]
5.29	270	−5.5	14	1	−106		−94	−130	SiGe HBT [21]
2.17	385	11.2	1.9	0.6	−120	−125	−122	−150	CMOS IC [22]
1.72	262	−11.5	75	1	−129		−111	−153	InGaP/GaAs HBT [23]
4.80	1200	4.8	36	1	−111		−95	−141	SiGe BiCMOS IC [24]
9.35	2500	18.3	570	1	−110		−82	−144	GaN/SiC pHEMT [25]
1.72	261	−10.3	55	1	−120		−103	−144	InGaP/GaAs HBT [26]
4.17	70	−6.1	102	1	−116		−96	−134	GaInP/GaAs HBT [27]
2.09	360		20.8	3	−140	−130	−117		CMOS VCO [28]
1.53	330		21.2	0.6	−133.5	−138	−125		CMOS VCO [29]
4.89	650		22	1	−124		−111		CMOS VCO [30]
1.85	280		20	3	−143	−133	−120		CMOS VCO [30]

which is also known as the **tuning gain**, K_0. This is the change in oscillation frequency for a change in control voltage. For the VCO in Figure 6-20,

$$K_0 = \frac{\Delta f_0}{\Delta v_i}. \tag{6.19}$$

The performance of a microwave VCO is one of the most competitive aspects of RF design, as every decibel reduction in phase noise greatly increases overall system performance. A high-performance VCO also relaxes demands on other system components. While FOM_1 (see Equation (6.18)) serves as a useful metric to compare VCOs, another FOM with bandwidth weighting provides a better comparison of the performance of different VCOs. This second figure of merit is [13]

$$\text{FOM}_2 = \mathcal{L}(f_m) - 10\log\left(\frac{1\text{ MHz}}{f_m}\right)^2 - 10\log\left(\frac{f_{\text{BW}}}{f_{\text{ref}}}\right), \tag{6.20}$$

where f_{BW} is the tuning bandwidth and f_{ref} is the reference bandwidth, taken here as 1 MHz. Again the phase noise is referenced to 1 MHz. A number of high-performance microwave oscillators are compared in Table 6-2. The best phase noise that can typically be achieved by VCOs operating in the 1–10 GHz range is -130 dBc/Hz at 1 MHz. This compares to the phase noise component of white noise at standard temperature, which was shown in Section 4.2.2 to be -177 dBc/Hz.

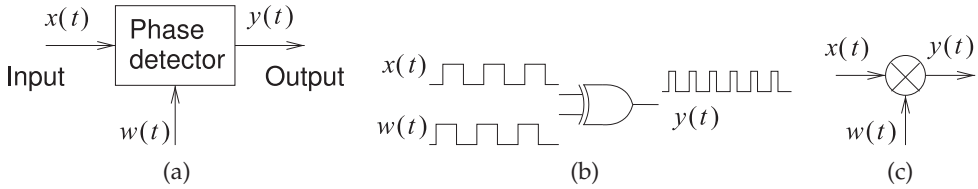


Figure 6-21: A phase detector: (a) block diagram; (b) a digital phase detector using an XOR gate; and (c) analog phase detector using a multiplier and filtering (not shown).

6.6 Phase Detector

A phase detector, also called a **phase comparator**, compares two waveforms and the output of the phase detector is a representation of the phase difference of the signals. There are two basic types of phase detectors: sinusoidal phase detectors and square signal phase detectors, which operate either in binary mode or by detecting zero crossings. The block diagram of a phase detector is shown in Figure 6-21(a) with the output $y(t)$ related to the difference of the phase of the input signals $x(t)$ and $w(t)$

A square wave detector is based on a logic circuit producing a signal that is averaged (or integrated) over time. An example is the XOR gate shown in Figure 6-21(b), which compares two digital signals that here have the same frequency but are shifted in phase. The output of the XOR gate, $y(t)$, is also a pulse train, and the average value of $y(t)$ is proportional to the phase difference of $x(t)$ and $w(t)$. This average value can be obtained by passing $y(t)$ through a lowpass filter to obtain a DC value. If $x(t)$ and $w(t)$ have different but close frequencies, the output of the lowpass filter will be a slowly varying signal. This circuit could be used to detect the phase difference of analog signals, but the signals first need to be converted to digital signals. This can be obtained through saturating amplification of the signals or by circuitry that responds to zero-crossings and produces a binary signal. Other digital phase comparators can be realized using charge-pumps, flip-flops, and sample-and-hold circuits.

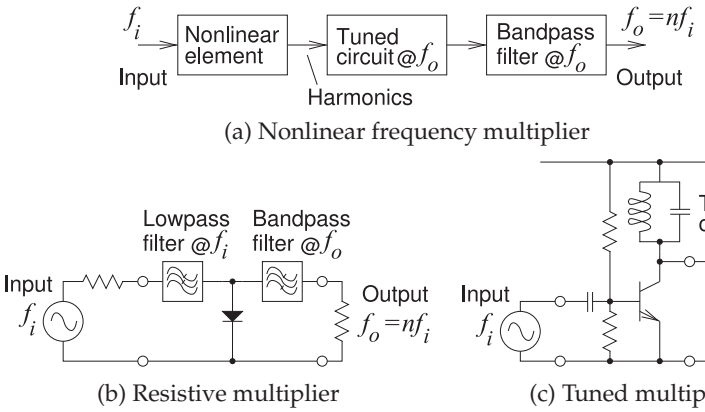
Sinusoidal phase detectors can use a mixer or an analog multiplier, as shown in Figure 6-21(c). The inputs could be two sinusoids with one of them adjusted in phase either externally or within the detector itself. So consider the inputs to the phase detector to be

$$x(t) = A_x \sin(\omega_x t + \phi_x) \quad \text{and} \quad w(t) = A_w \cos(\omega_w t + \phi_w). \tag{6.21}$$

Then the output of the multiplier is

$$\begin{aligned} y(t) &= A_x \sin(\omega_x t + \phi_x) A_w \cos(\omega_w t + \phi_w) \\ &= \tfrac{1}{2} A_x A_w \left[\sin(\omega_x t + \omega_w t + \phi_x + \phi_w) + \sin(\omega_x t - \omega_w t + \phi_x - \phi_w)\right]. \end{aligned} \tag{6.22}$$

Typically the frequencies of the two input signals are close so that $\omega_x = \omega + \frac{1}{2}\Delta\omega$ and $\omega_w = \omega - \frac{1}{2}\Delta\omega$, where $\Delta\omega$ is small and in most phase detector applications either $\Delta\omega = 0$ or a feedback loop attempts to set $\Delta\omega$ to zero so there is little frequency error. So, following lowpass filtering, the output of



(a) Nonlinear frequency multiplier

(b) Resistive multiplier

(c) Tuned multiplier

Figure 6-22: Frequency multipliers.

the phase detector is

$$y(t) = \tfrac{1}{2} A_x A_w \sin(\Delta\omega t + \phi_x - \phi_w). \tag{6.23}$$

Now $\Delta\omega t$ can be considered to be a phase error and incorporated in effective phases of the input signals, $\psi_x(t) = \frac{1}{2}\Delta\omega t + \phi_x$ and $\psi_w(t) = -\frac{1}{2}\Delta\omega t + \phi_w$. Then

$$y(t) = \tfrac{1}{2} A_x A_w \sin[\psi_x(t) - \psi_w(t)]. \tag{6.24}$$

This is a functional phase detector provided that the difference in the phases of the input signals is between $-\pi/2$ and $\pi/2$.

A variation of the sinusoidal detector is called a **phase-frequency detector** (**PFD**) which has an extended range and the effective phase difference of the input signals can have a magnitude greater than $\pi/2$ [31, 32].

The output signal, $y(t)$, is proportional to the amplitudes of the input signals and to their phase difference. In practice the amplitudes of both input signals are scaled to a constant amplitude so that the output only depends on the phase difference.

When a mixer is used in a phase detector, it operates much the same way as a multiplier. With a switching mixer one of the input signals becomes the drive of the switching part of the mixer and the other signal is the sinusoidal drive. Now only the second input need be scaled to have a constant amplitude.

6.7 Frequency Multiplier

There are three main types of microwave frequency multipliers. One uses a nonlinear resistive element to generate harmonics and a bandpass filter selects the appropriate harmonic for the output (see Figure 6-22(a)) [33]. A second type uses a reactive element but is inherently narrow band. A third type uses a mixer.

An example is the diode frequency multiplier shown in Figure 6-22(b). If unbiased, the diode conducts current only during the positive half-cycle of the input signal, creating a voltage across the diode that is rich in harmonics. If no external bias is applied, some of the positive voltage swing is required to turn the diode on. This type of multiplier is referred to as a **resistive frequency multiplier**. The transistor circuit in Figure 6-22(c) uses the resistive multiplier principle to achieve frequency multiplication. Now

the bandpass filter is the tuned circuit in the collector leg of the amplifier. The transistor is biased to realize a strong nonlinearity and the base-emitter junction has an exponential current-voltage characteristic.

If the input signal is

$$x(t) = A\cos(\omega_t + \phi) \tag{6.25}$$

then the output at the n harmonic is

$$y(t) = A_n \cos(n\omega_t + n\phi). \tag{6.26}$$

One of the issues with all frequency multipliers is the multiplication of the phase noise on the input signal. If ϕ represents the phase noise on the original signal, then the phase noise of the output signal will be increased by a factor n. That is, even if the frequency multiplier introduces no noise of its own, the signal-to-noise (i.e., phase noise) ratio of a signal will be reduced by a factor n. The amplitude of the output signal, A_n, will usually be much less than that of the input signal unless the nonlinear circuit incorporates an amplifier. The increase in phase noise and power loss (without amplification) are the major drawbacks of using a resistive nonlinear element to produce frequency multiplication.

The bandwidth of a resistive frequency multiplier is limited by the maximum bandwidth of a filter that will select just one harmonic at the output. That is, if the input signal is at 10 GHz and with $10\times$ multiplication the output will be at 100 GHz. The useful fractional input bandwidth is $1/n = \frac{1}{10}$ of the input signal, otherwise two harmonics could appear simultaneously in the output.

Another type of frequency multiplier uses a reactive element such as the nonlinear capacitance of a reverse-biased semiconductor diode (i.e., a varactor diode) [34]. This mixer is called a **reactive frequency multiplier** or a **parametric frequency multiplier**. The nonlinear reactive element is part of a resonant input circuit at the input frequency and also part of a resonant output circuit at the output frequency. The efficiency of this type of mixer is higher than for a resistive mixer but the bandwidth is much lower.

A third type of frequency multiplier uses a mixer. Some, but not all, mixers can be used to realize frequency multiplication by applying the same signal to the two input ports. The suitable mixer circuits are the balanced mixers but not the mixers that rely on filtering to separate LO, RF, and IF signals.

A further type of frequency multiplier uses a phase-locked loop. Frequency division in the feedback loop results in frequency multiplication of the input signal. Yet another type uses flip-flops working on a microwave frequency binary clock signal and this is the type often used in modern RFICs to provide a square wave drive to a switching mixer.

6.8 Frequency Divider

A frequency divider is a module that reduces the frequency of a signal. There are three main types of frequency dividers: those that work with square waves and those that work with sinusoidal signals. The square wave dividers are much simpler. A divide-by-2 square wave divider is shown in Figure 6-23. The square wave input can be produced from a sinusoidal signal using zero-crossing detection or a high-gain circuit with saturating levels.

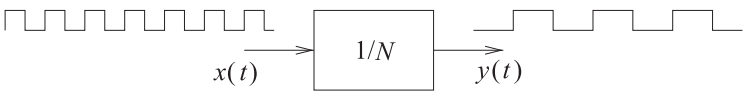


Figure 6-23: Digital frequency divider with the waveforms shown for a divide-by-2 divider.

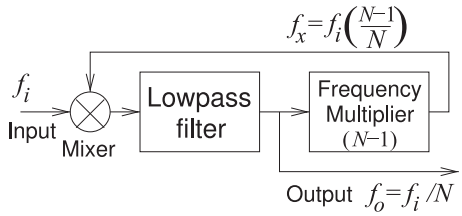


Figure 6-24: Regenerative frequency divider.

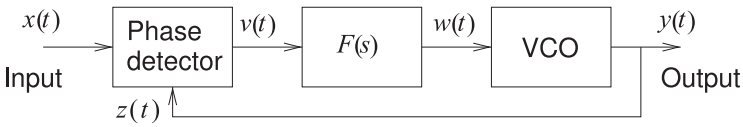


Figure 6-25: A phase-locked loop with a phase-detector and a frequency divider indicated by $(1/N)$.

One type of analog frequency divider is the regenerative frequency divider shown in Figure 6-24. The key element of this circuit is the mixer, which here produces an output at the difference frequency of the input at frequency f_i and the signal is fed back at frequency f_x. The output of the mixer is the lowpass filtered difference of f_i and f_x. The loop stabilizes to produce the divided frequency at the output.

Another analog frequency divider, called a **locked-oscillator frequency divider**, uses injection locking of an oscillator [35]. It is relatively easy to lock many oscillators by injecting a signal near the nth harmonic of the free-running oscillation frequency. Then the oscillation frequency shifts and the output has a frequency $1/n$th that of the input signal.

Yet another type uses flip-flops dividing the frequency of a microwave-frequency binary clock signal.

6.9 Phase-Locked Loop (PLL)

A phase-locked loop (PLL) is a feedback system in which the frequency and phase of an output signal is related to the frequency and phase of an input signal. The block diagram of a PLL is shown in Figure 6-25. An input signal $x(t)$ is compared to a feedback signal $z(t)$. The frequency of $y(t)$ will be the average frequency of $x(t)$. The way the loop achieves this is that the output of the phase detector is proportional to the phase difference of $x(t)$ and $z(t)$. This is then filtered by the block $F(s)$ (generally a lowpass filter) to produce a DC-like signal that drives a VCO. The filtering block, $F(s)$, removes undesired components from the phase detector and also sets the dynamic response of the PLL.

6.9.1 Operation

The operation of the PLL in Figure 6-25 can be modeled as a linear system with the assumption that $x(t)$ and $z(t)$ (= $y(t)$ here) are nearly periodic signals. Thus, approximately,

$$x(t) = A_x \cos(\omega_x t + \phi_x) \quad \text{and} \quad z(t) = A_w \cos(\omega_z t + \phi_z). \tag{6.27}$$

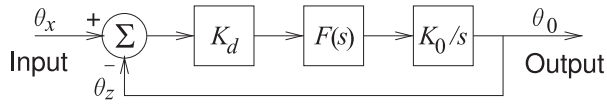


Figure 6-26: Linearized model of an analog phase-locked loop.

PLLs require that the radian frequency ω_x be close to ω_z, which is near the free-running frequency of the VCO. (In practice, ω_x is usually much less than ω_z and a frequency divider is used to divide the signal at ω_z to obtain a signal with a frequency close to ω_x. The analysis is similar to that presented here.) This defines the **capture range** of the PLL. Therefore Equation (6.27) can be written

$$x(t) = A_x \cos(\Theta_x(t)) \quad \text{and} \quad z(t) = A_w \cos(\Theta_0(t)). \tag{6.28}$$

The phases Θ_x and Θ_0 incorporate the original phases ϕ_x and ϕ_w, respectively, and the effective time-dependent phase difference due to the small time-dependent difference of the frequencies of $x(t)$ and $z(t)$. Analysis of the PLL in Figure 6-25 begins with the voltage at the output of the phase detector, $v(t)$, which is proportional to the phase difference of the two input signals and independent of their amplitude:

$$v(t) = K_d(\Theta_x - \Theta_0), \tag{6.29}$$

where K_d is the **phase detector gain factor**. The output of the phase detector is filtered by the block with transfer function $F(s)$. Usually this block is a lowpass filter, but there are applications where it could be a bandpass filter or have some other characteristic.

The output of the VCO is controlled by the voltage $w(t)$ producing a signal with frequency

$$f_0 = f_c + \Delta f = f_c + K_0 v(t), \tag{6.30}$$

where f_c is the frequency of oscillation when the control voltage is zero and K_0 is the VCO gain factor.

Equations (6.29) and (6.30) describe the linear system shown in Figure 6-26. The transfer function of this system is

$$\frac{\theta_0}{\theta_x} = \frac{K_d K_0 F(s)/s}{1 + K_d K_0 F(s)/s} = \frac{G(s)}{1 + G(s)}, \tag{6.31}$$

where $G(s) = K_d K_0 F(s)$. The phase error function is

$$\epsilon(s) = \theta_x(s) - \theta_0 = \theta_x \left[1 - \frac{K_d K_0 F(s)}{s + K_d K_0 F(s)}\right] = \frac{s\theta_x(s)}{s + K_d K_0 F(s)}. \tag{6.32}$$

So the phase error function is directly related to the phase of the input signal $x(t)$. In the next section a particular choice of the transfer function $F(s)$ is used, leading to the identification of a particular PLL application.

6.9.2 First-Order PLL

Without a filter (i.e., $F(s) = 1$) the PLL is referred to as a first-order PLL or **first-order loop** and its transfer function is

$$\frac{\theta_0(s)}{\theta_x(0)} = \frac{K_d K_0/s}{1 + K_d K_0/s} = \frac{K_d K_0}{s + K_d K_0}. \tag{6.33}$$

This seems to be the transfer function of a lowpass filter, but the input to the system is phase and the output is the phase deviation from the free-running frequency of the VCO. The phase error function of the first-order PLL is

$$\epsilon(s) = \theta_x(s) - \theta_0 = \theta_x(s)\left(1 - \frac{K_d K_0}{s + K_d K_0}\right) = \frac{s\theta_x(s)}{s + K_d K_0}. \tag{6.34}$$

The steady-state response of the first-order PLL to specific input signals will now be derived. The long-term (steady-state) behavior is developed using the **final-value theorem**. A function $H(s)$ with all poles in the left-half plane has a time-domain response $h(t)$ with long-term behavior given by

$$\lim_{t\to\infty} h(t) = \lim_{s\to 0} sH(s). \tag{6.35}$$

Consider now the response to a step change, $\Delta\theta_x$, of the phase of the input signal. The Laplace transform of a step function of magnitude $\Delta\theta_x$ is $\Delta\theta_x/s$ and this is what is used in Equation (6.34). Thus the steady-state error function, $\epsilon_{ss} = \epsilon(t)$ as $t \to \infty$, is

$$\begin{aligned}\epsilon_{ss} &= \lim_{s\to 0}\left[\frac{s^2\theta_x(s)}{s + K_d K_0}\right] = \lim_{s\to 0}\left[\frac{s^2\Delta\theta_x/s}{s + K_d K_0}\right] = \lim_{s\to 0}\left[\frac{s\Delta\theta_x}{s + K_d K_0}\right] \\ &= 0.\end{aligned} \tag{6.36}$$

Equation (6.36) indicates that a first-order PLL will eventually track a phase change of the input signal. However, it will not respond immediately as the loop has a lowpass characteristic. In effect the PLL is acting as a high-Q bandpass filter.

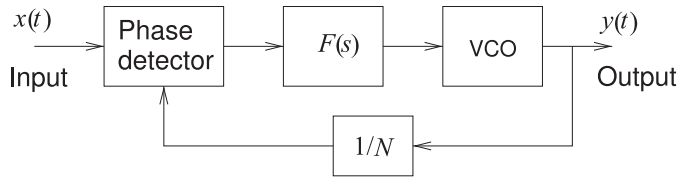
Now consider a step change in the frequency of the input signal. If the step change in the radian frequency of $x(t)$ is $\Delta\omega_x$, then the resulting phase change will be a ramp so that, in the Laplace domain, $\theta(s) = \Delta\omega_x/s^2$. Then the steady-state error is

$$\begin{aligned}\epsilon_{ss} &= \lim_{s\to 0}\left[\frac{s^2\Delta\omega_x/s^2}{s + K_d K_0}\right] = \lim_{s\to 0}\left[\frac{\Delta\omega_x}{s + K_d K_0}\right] \\ &= \Delta\omega_x/(K_d K_0).\end{aligned} \tag{6.37}$$

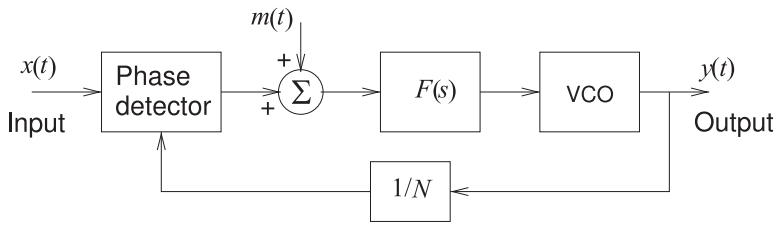
Equation (6.37) indicates that a first-order PLL will follow a change in frequency with a phase error that is proportional to the magnitude of the frequency change. So if the input signal $x(t)$ is a frequency modulated signal, the first-order PLL will demodulate the signal and recover the original modulation as the error signal. Thus a first-order PLL demodulates an FM signal, that is, it is a **frequency demodulator**.

6.9.3 Applications

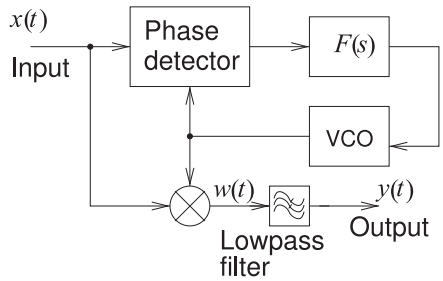
Three applications of the PLL are shown in Figure 6-27. A **frequency synthesizer** is used to create a signal locked to a fixed-frequency and very accurate reference oscillator but at another frequency (see Figure 6-27(a)). Typically the reference is a precision low-frequency reference oscillator, such as a quartz crystal oscillator [36]. The output frequency is not at the same frequency as the reference oscillator. This is accomplished by including a frequency divider in the PLL. Normally the division factor, N, is an integer



(a) Frequency synthesizer



(b) Phase modulator



(c) Coherent amplitude demodulator

Figure 6-27: Applications using phase-locked loops.

so that the output signal's frequency will be an integer multiple of the input frequency. A fractional-N frequency synthesizer is achieved using two integer dividers with division factors N and M. A controller alternates between the two division factors so that the VCO tends to output first one frequency and then the next. The VCO stabilizes at a frequency that is the weighted average of the two division factors. If the frequency is divided by M for a time τ_M, then by N for time τ_N, and repeated, then the effective division factor is $(\tau_M M + \tau_N N)/(\tau_M + \tau_N)$. The loop dynamics (determined by the filter $F(s)$) is chosen so that the VCO can only change more slowly than the toggling of the division factors. The division factors, and the toggling between its two values, are chosen to synthesize an output signal with the desired frequency.

The second application of the PLL is a phase modulator, see Figure 6-27(b). The phase modulating signal $m(t)$ effectively is adding to the phase error generated by the phase detector. This results in a change in the phase of the VCO output compensating for the inserted phase change.

The third application of the PLL is the **coherent amplitude demodulator** shown in Figure 6-27(c). The input is an amplitude modulated signal

$$x(t) = A\,[1 + m(t)]\sin(\omega_0 t), \tag{6.38}$$

which is demodulated to recover $m(t)$ by multiplying $x(t)$ by an LO signal with the same carrier frequency. The characteristic of $F(s)$ is chosen so that the PLL generates the LO signal as the loop establishes the oscillation

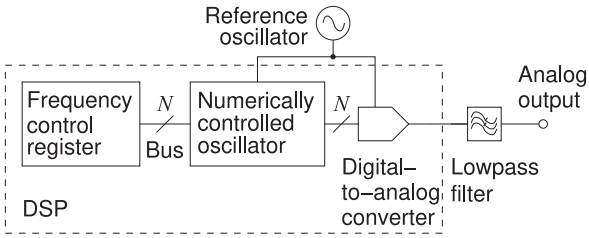


Figure 6-28: Direct digital synthesis (DDS) module.

frequency as the central frequency of the modulated input. The loop filter $F(s)$ ensures that the VCO frequency cannot change too quickly and the output frequency of the VCO approximates a sinusoidal signal having a radian frequency ω_0 with a phase offset. Thus

$$w(t) = x(t)\sin(\omega_0 t + \theta) = A\left[1 + m(t)\right]\tfrac{1}{2}\left[\cos\theta - \cos(2\omega_0 t_\theta)\right]. \tag{6.39}$$

Following lowpass filtering the output is

$$y(t) = A\left[1 + m(t)\right]\cos\theta. \tag{6.40}$$

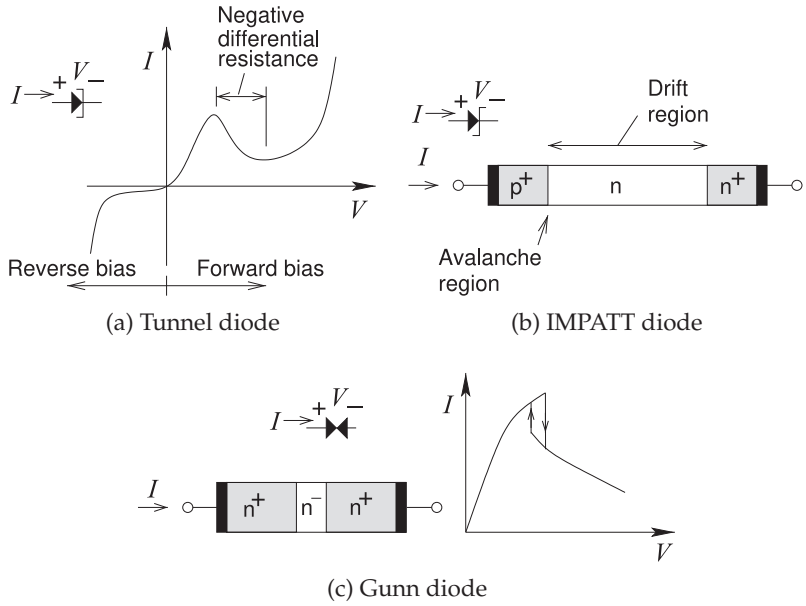
With loop dynamics chosen so that θ is small, the original modulation, $m(t)$, is recovered. This AM demodulator performs well when the input SNR is low as demodulation is coherent.

6.10 Direct Digital Synthesizer

A direct digital synthesizer (DDS) uses a DSP unit, commonly called a **numerically controlled oscillator**, to create the digital version of a modulated waveform (see Figure 6-28). The digital output of the DSP drives a DAC that is filtered by a lowpass filter producing the analog output. A DDS module enables a fairly complex RF system to be realized using just a few components, as frequency generation and frequency synthesis are embodied in one unit. The accuracy of the frequency of the analog signal is established by the reference oscillator that acts as a precision clock for the DSP and DAC.

The numerically controlled oscillator has an accumulator, often called a **phase accumulator**, that counts out the phase of the analog signal to be synthesized. If there is modulation, then it is incorporated at this stage. Both modulation and the center frequency of the modulated output are changed by changing the accumulator's step size. The accumulator drives a sine look-up table that then produces the digital words presented to the DAC.

A DDS has the advantage that the output frequency can be changed almost instantaneously which is in contrast to a PLL in which the rate of frequency change of the output is limited by the bandwidth of the loop filter. A major disadvantage of a DDS is the number of spurious signals generated by quantization issues, and this is mostly related to the bit length of the accumulator. There are also aliasing or image signals generated as images of the intended signal relative to the clock frequency. However, the spurious signals from two DDS units with the same reference oscillator are correlated and this can be exploited in some system architectures to remove the effect of spurious signals. Most of the negative characteristics of a DDS are reduced by using a longer accumulator and hence a DAC with a larger number of input bits. This, however, leads to increased power consumption. These drawbacks are sometimes insignificant when compared to the frequency agility and system simplicity achieved.



(a) Tunnel diode (b) IMPATT diode

(c) Gunn diode

Figure 6-29: Two terminal semiconductor sources

6.11 Diode and Vacuum Sources

There are many types of sources based on special properties of diodes or of vacuum devices. This section describes some of these sources. While sources based on transistor circuits are preferred for integration and manufacturability, diode and vacuum sources are in widespread use because of either the high power that can be efficiently generated or because they operate at high frequencies beyond the reach of transistor circuits.

6.11.1 Two-Terminal Semiconductor Sources

Two-terminal semiconductor sources are based on exploiting the negative resistance presented at the terminals of several types of devices [37–40]. They either exploit a dynamic negative resistance created in a high field region of a semiconductor due to a charge imbalance, or they exploit the finite time for carriers to cross a semiconductor region and thus produce an RF current that is out of phase with the applied voltage, thus presenting a negative RF resistance. Three representative devices are the tunnel diode, the IMPATT diode, and the Gunn diode.

Tunnel Diode

A tunnel diode is a pn junction diode in which the conduction band states on the n side are filled with electrons and these line up with empty valance band states (i.e. holes) on the p side [37–46]. This results in a very narrow pn junction barrier. For negative voltages and small and large applied voltages the diode acts as a conventional pn junction diode with an exponential current-voltage characteristic. However, as the applied voltages increase above zero, the conduction and valance bands become more misaligned and the voltage eventually drops before increasing again. This effect is due to quantum mechanical tunnelling. This is seen in the current-voltage characteristic of the tunnel diode shown in Figure 6-29(a), where the drop in

current creates a negative dynamic resistance. Embedded in an appropriate circuit, even as simple as a parallel RLC circuit, the tunnel diode is the active element of an RF oscillator. Usually a Gunn diode has a very high dopant concentration so that the reverse breakdown voltage is very low.

IMPATT Diode

An IMPATT diode (IMPact ionization Avalanche Transit-Time) produces high power and throughout the history of its use it has produced the highest power levels at the highest frequency, up to 1000 GHz, of any semiconductor device [37–40, 47–51].

The IMPATT diode is the most important of the transit time semiconductor diodes, see Figure 6-29(b). In transit time devices the generation of charge carriers is concentrated in one narrow region of the diode. In an IMPATT diode, a high field at the boundary between a highly doped n region, the n^+ region, and a lightly doped n region leads to avalanche multiplication producing holes and electrons. The holes are quickly collected by an adjacent metal contact and the electrons transit through a drift region with usually intrinsic doping and a constant field. If the drifting electrons are sufficiently delayed, then the RF current through the device will be out of phase with the applied RF voltage (superimposed on a biasing DC voltage) and a negative RF resistance is presented at the device terminals. The roles of the holes and electrons can also be exchanged. Thus the IMPATT diode can be used as the active component of an oscillator [52]. They can be used in an amplifier as a reflection device having a reflection coefficient greater than one [53]. However, the oscillating signal produced has high noise due to the underlying avalanche process.

Other effects can produce charges that eventually drift and produce a negative RF resistance. An example is a **TUNNETT** diode that injects charges through tunneling [54, 55]. Another device is the **TRAPATT** diode (trapped plasma avalanche transit time diode), a pn junction diode, where the carrier injection results from a trapped space-charge plasma formed within the junction region [42, 56].

Gunn Diode

A Gunn diode is also called a transferred electron device or a Gunn effect device. While strictly not a diode as there is not a junction, the name Gunn diode has become common usage because there are two electrodes. The structure and current-voltage characteristic of a Gunn diode are shown in Figure 6-29(c) [37–40, 57–61]. The device has three n-type regions: two heavily doped n^+ regions at each contact, separated by a lightly doped n^- region. When a voltage is applied to the device, most of the voltage is across the n^- region and the device acts like a resistor with the current through the Gunn diode proportional to the voltage across it. At higher voltages the conductivity of the n^- region drops and the current drops, so that there is a region of negative dynamic resistance.

6.11.2 Vacuum Devices

Some microwave systems use vacuum tubes to obtain high RF powers; the most common vacuum tube devices are reviewed here.

Magnetron

The magnetron is the original device used for generating microwave power and was invented during World War II for use in radar equipment. It is most commonly used in microwave ovens, where it is the most efficient means of producing microwave power at 2.4 GHz. It is used in military systems today to produce megawatts of pulsed RF power and is used to generate substantial power up to a few terahertz [62–65]. In a magnetron a circular chamber, containing the cathode, is surrounded by and connected to a number of resonant cavities. The walls of the chamber are the anode. The cavity dimensions determine the frequency of the output signal. A strong constant magnetic field in the chamber causes electrons that want to flow from the cathode to the anode to rotate. As the electrons pass the entrance of the circular cavities, the electrons interact with the EM field in the cavity, enhancing the field at a characteristic frequency, which is the frequency of oscillation of the magnetron.

Klystron

The klystron is a long, narrow vacuum tube with an electron gun (the cathode) at one end and an anode at the other [66–69]. In between is a series of doughnut-shaped resonant cavities aligned so that the electron beam from the cathode passes through the hole. As the beam passes the cavities, small changes in the electron beam affect the EM field in the cavities. The EM fields in the cavities begin to oscillate, which in turn affect the passing electron beam. A feedback effect results, and when the last and first resonant cavities are connected, a large oscillating microwave signal is produced.

Traveling Wave Tube and Backward Wave Oscillator

A traveling wave tube (TWT) is a long vacuum tube with an electron gun (the cathode) at one end and a collector (the anode) at the other. The electron beam produced by the gun travels down the center of a long wire helix [70–75]. An excitation coupling probe, an antenna, placed near the cathode introduces a microwave signal. This modulates the electron beam, which in turn induces an RF current in the helix. The helix is designed so that the EM signal supported by the helix travels at the same speed as the electron beam so that the electron beam and EM signal guided by the helix continuously interact. Thus the RF current in the helix as well as the RF field grow in strength along the tube. An output probe at the anode end of the tube couples to the RF field and delivers an amplified version of the input RF signal. As an amplifier the TWT is called a traveling wave tube amplifier (TWTA).

In the TWT the helix slows the RF signal down to match the speed of the electrons and so the helix is called a slow-wave structure. Another tube device that uses a slow-wave structure is the backward wave oscillator (BWO). However, in the BWO the beam is directed against the traveling wave supported by the helix.

6.12 Summary

This chapter described a number of complex mixer and oscillator modules that are available to assemble high-performance RF and microwave systems. These modules and the other modules used in this book are often available as off-the-shelf items from companies that specialize in designing modules. This is a cost effective way of developing low and medium volume microwave systems such as basestation hardware. For a competitive advantage a company selling a system or subsystem may develop some of the modules themselves. For a high volume market, say for a cell phone, many of the modules would be integrated on-chip. Design is then an expensive proposition but overall performance and size would be optimized.

6.13 References

[1] S. Maas, *Microwave Mixers*. Artech House, 1986.

[2] M. Steer and P. Khan, "An algebraic formula for the output of a system with large-signal, multifrequency excitation," *Proc. of the IEEE*, vol. 71, no. 1, pp. 177–179, 1983.

[3] R. Pettai, *Noise in Receiving Systems*. John Wiley & Sons, 1996.

[4] N. Carvalho, J. Pedro, W. Jang, and M. Steer, "Nonlinear simulation of mixers for assessing system-level performance," *Int. J. Microwave Millimeter Wave Computer Aided Engineering*, vol. 15, no. 4, pp. 350–361, Jul. 2005.

[5] B. Gilbert, "A precise four-quadrant multiplier with subnanosecond response," *IEEE J. Solid-State Circuits*, vol. 3, no. 4, pp. 365–373, Dec. 1968.

[6] X. Yang, A. Davierwalla, D. Mann, and K. Gard, "A 90nm CMOS direct conversion transmitter for WCDMA," in *2007 IEEE Radio Frequency Integrated Circuits (RFIC) Symp.*, Jun. 2007, pp. 17–20.

[7] X. Yang, "90nm cmos transmitter design for WCDMA," Ph.D. dissertation, North Carolina State University, 2009.

[8] M. Ding, K. Gard, and M. Steer, "A highly linear and efficient CMOS RF power amplifier with a new circuit synthesis technique," *IEEE Trans. on Microwave Theory and Techniques*, vol. 60, no. 8, pp. 1–2, Nov. 2012.

[9] M. Steer, *Microwave and RF Design, Amplifiers and Oscillators*, 3rd ed. North Carolina State University, 2019.

[10] D. Leeson, "A simple model of feedback oscillator noise spectrum," *Proc. of the IEEE*, vol. 54, no. 2, pp. 329–330, Feb. 1966.

[11] A. Hajimiri and T. Lee, "Design issues in cmos differential lc oscillators," *IEEE J. of Solid-State Circuits*, vol. 34, no. 5, pp. 717–724, May 1999.

[12] P. Kinget, *Integrated GHz Voltage Controlled Oscillators*. Kluwer, 1999.

[13] A. Victor and M. Steer, "Reflection coefficient shaping of a 5-GHz voltage-tuned oscillator for improved tuning," *IEEE Trans. on Microwave Theory and Techniques*, vol. 55, no. 12, pp. 2488–2494, Dec. 2007.

[14] S.-S. Myoung and J.-G. Yook, "Low-phase-noise high-efficiency MMIC VCO based on InGaP/GaAs HBT with the LC filter," *Microwave and Optical Technology Letters*, vol. 44, no. 1, pp. 123–126, Jan. 2005.

[15] C.-H. Lee, S. Han, B. Matinpour, and J. Laskar, "A low phase noise X-band MMIC GaAs MESFET VCO," *IEEE Microwave and Guided Wave Letters*, vol. 10, no. 8, pp. 325–327, Aug. 2000.

[16] Z. Cheng, Y. Cai, J. Liu, Y. Zhou, K. Lau, and K. Chen, "A low phase-noise X-band MMIC VCO using high-linearity and low-noise composite-channel $Al_{0.3}Ga_{0.7}N/Al_{0.05}Ga_{0.95}N/GaN$ hemts," *IEEE Trans. on Microwave Theory and Techniques*, vol. 55, no. 1, pp. 23–29, Jan. 2007.

[17] H. Zirath, R. Kozhuharov, and M. Ferndahl, "Balanced colpitt oscillator mmics designed for ultra-low phase noise," *IEEE J. of Solid-State Circuits*, vol. 40, no. 10, pp. 2077–2086, Oct. 2005.

[18] Y.-K. Chu and H.-R. Chuang, "A fully integrated 5.8-GHz U-NII band 0.18–μ;m CMOS VCO," *IEEE Microwave and Wireless Components Letters*, vol. 13, no. 7, pp. 287–289, Jul. 2003.

[19] C. Meng, Y. Chang, and S. Tseng, "4.9-GHz low-phase-noise transformer-based superharmonic-coupled GaInP/GaAs HBT QVCO," *IEEE Microwave and Wireless Components Letters*, vol. 16, no. 6, pp. 339–341, Jun. 2006.

[20] C. Meng, C. Chen, Y. Chang, and G. Huang, "5.4 ghz-127 dbc/hz at 1 MHz GaInP/GaAs

HBT quadrature VCO using stacked transformers," *Electronics Letters*, vol. 41, no. 16, pp. 33–34, Aug. 2005.

[21] T. Hancock and G. Rebeiz, "A novel superharmonic coupling topology for quadrature oscillator design at 6 ghz," in *2004 IEEE Radio Frequency Integrated Circuits (RFIC) Symp. Digest of Papers*, Jun. 2004, pp. 285–288.

[22] S.-W. Yoon, S. Pinel, and J. Laskar, "A 0.35-μm CMOS 2-GHz VCO in wafer-level package," *IEEE Microwave and Wireless Components Letters*, vol. 15, no. 4, pp. 229–231, Apr. 2005.

[23] J.-H. Yoon, S.-H. Lee, A.-R. Koh, G. Kennedy, P. Gary, and N.-Y. Kim, "Optimized phase noise of LC VCO using an asymmetric-inductance tank in InGaP/GaAs HBT technology," *Microwave and Optical Technology Letters*, vol. 48, no. 6, pp. 1035–1040, Jun. 2006.

[24] O. Esame, I. Tekin, A. Bozkurt, and Y. Gurbuz, "Design of a 4.2–5.4 GHz differential LC VCO using 0.35 μm SiGe BiCMOS technology for ieee 802.11a applications," *Int. J. of RF and Microwave Computer-Aided Engineering*, vol. 17, no. 2, pp. 243–251, 2007.

[25] A. Maas and F. van Vliet, "A low-noise X-band microstrip VCO with 2.5 GHz tuning range using a GaN-on-SiC p-HEMT," in *2005 European Gallium Arsenide and Other Semiconductor Application Symp., (EGAAS 2005)*, Oct. 2005, pp. 1805–1808.

[26] J.-H. Yoon, S.-H. Lee, A.-R. Koh, B. Shrestha, S.-H. Cheon, G. Kennedy, and N.-Y. Kim, "A novel harmonic noise frequency filtering VCO for optimizing phase noise," in *2006 IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 2006, pp. 1805–1808.

[27] C. Meng, S. Tseng, Y. Chang, J. Su, and G. Huang, "4-GHz low-phase-noise transformer-based top-series GaInP/GaAs HBT QVCO," in *2006 IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 2006, pp. 1809–1812.

[28] P. Andreani and X. Wang, "On the phase-noise and phase-error performances of multiphase lc cmos vcos," *IEEE J. of Solid-State Circuits*, vol. 39, no. 11, pp. 1883–1893, Nov. 2004.

[29] P. Vancorenland and M. Steyaert, "A 1.57 GHz fully integrated very low phase noise quadrature vco," in *2001 Symp. on VLSI Circuits Dig. of Technical Papers*, 2001, pp. 111–114.

[30] S. Gierkink, S. Levantino, R. Frye, and V. Boccuzzi, "A low-phase-noise 5 GHz quadrature CMOS VCO using common-mode inductive coupling," in *Proc. of the 28th European Solid-State Circuits Conf., (ESSCIRC 2002)*, Sep. 2002, pp. 539–542.

[31] J. Savoj and B. Razavi, "A 10-gb/s cmos clock and data recovery circuit with a half-rate binary phase/frequency detector," *IEEE J. of Solid-State Circuits*, vol. 38, no. 1, pp. 13–21, Jan. 2003.

[32] H. Johansson, "A simple precharged cmos phase frequency detector," *IEEE J. of Solid-State Circuits*, vol. 33, no. 2, pp. 295–299, Feb. 1998.

[33] A. Raisanen, "Frequency multipliers for millimeter and submillimeter wavelengths," *Proc. of the IEEE*, vol. 80, no. 11, pp. 1842–1852, Nov. 1992.

[34] D. Holcomb, "Parametric frequency multiplier," US Patent US Patent 3 076 133, Jan. 29, 1963.

[35] M. Tiebout, "A cmos direct injection-locked oscillator topology as high-frequency low-power frequency divider," *IEEE J. of Solid-State Circuits*, vol. 39, no. 7, pp. 1170–1174, Jul. 2004.

[36] M. Odyniec, Ed., *RF and Microwave Oscillator Design*. Artech House, 2002.

[37] M. Shur, *GaAs Devices and Circuits*. Springer, 1987.

[38] K. Ng, *Complete Guide to Semiconductor Devices*. Wiley, 2002.

[39] U. Mishra and J. Singh, *Semiconductor Devices: Physics and Technology*. John Wiley & Sons, 2008.

[40] ——, *Semiconductor Device Physics and Design*. Springer, 2007.

[41] B. B. Streetman and S. Banerjee, *Solid State Electronic Devices*, 6th ed. Prentice Hall, 2006.

[42] S. Sze and K. Ng, *Physics of Semiconductor Devices*, 3rd ed. John Wiley & Sons, 2007.

[43] W. Chow, *Principles of Tunnel Diode Circuits*. Wiley, 1964.

[44] L. Esaki, "Discovery of the tunnel diode," *IEEE Trans. on Electron Devices*, vol. 23, no. 7, pp. 644–647, 1976.

[45] K. Ng, *Complete Guide to Semiconductor Devices*. Wiley, 2002, ch. Tunnel diode.

[46] D. Schroder, *Semiconductor Material and Device Characterization*. IEEE Press and Wiley, 2006.

[47] T. Midford and R. Bernick, "Millimeter-wave cw impatt diodes and oscillators," *IEEE Trans. on Microwave Theory and Techniques*, vol. 27, no. 5, pp. 483–492, May 1979.

[48] H. Xu, M. Morschbach, J. Werner, and E. Kasper, "Monolithic integrated oscillator with silicon impatt diode for automotive radar applications," in *European Microwave Integrated Circuit Conference, 2008*, Oct. 2008, pp. 20–23.

[49] P. Mukherjee and B. Gupta, "Terahertz (thz) frequency sources and antennas—a brief review," *Int. J. Infrared and Millimeter Waves*, vol. 29, no. 12, pp. 1091–1102, 2008.

[50] M.-S. Gupta, R. Lomax, and G. Haddad, "Noise considerations in self-mixing impatt-diode oscillators for short-range doppler radar applications," *IEEE Trans. on Microwave Theory and Techniques*, vol. 22, no. 1, pp. 37–43, Jan. 1974.

[51] W. Evans and G. Haddad, "A large signal analysis of impatt diodes," *IEEE Trans. on Electron Devices*, vol. 15, no. 10, pp. 708–717, Oct. 1968.

[52] K. Kurokawa and F. Magalhaes, "An X-band 10-watt multiple-impatt oscillator," *Proc. of the IEEE*, vol. 59, no. 1, pp. 102–103, Jan. 1971.

[53] M.-S. Gupta, "Large-signal equivalent circuit for impatt-diode characterization and its application to amplifiers," *IEEE Trans. on Microwave Theory and Techniques*, vol. 21, no. 11, pp. 689–694, Nov. 1973.

[54] H. Eisele, A. Rydberg, and G. Haddad, "Recent advances in the performance of InP Gunn devices and GaAs TUNNETT diodes for the 100-300-GHz frequency range and above," *IEEE Trans. on Microwave Theory and Techniques*, vol. 48, no. 4, pp. 626–631, Apr. 2000.

[55] J.-I. Nishizawa, T. Kurabayashi, Y. Miura, T. Sawai, P. Plotka, and M. Watenabe, "Development of sub-thz tunnett diode for biomedical imaging," in *Infrared and Millimeter Waves, 2007 and the 2007 15th International Conference on Terahertz Electronics. IRMMW-THz. Joint 32nd International Conference on*, Sep. 2007, pp. 150–151.

[56] J. DeLoach, B.C. and D. Scharfetter, "Device physics of trapatt oscillators," *IEEE Trans. on Electron Devices*, vol. 17, no. 1, pp. 9–21, Jan. 1970.

[57] B. Bosch and R. Engelmann, *The Gunn Effect*. Clarendon Press, 1974.

[58] G. Hobson, *The Gunn Effect*. Clarendon Press, 1974.

[59] H. Eisele, "480 GHz oscillator with an InP Gunn device," *Electronics Letters*, vol. 46, no. 6, pp. 422–423, Mar. 2010.

[60] F. Amir, C. Mitchell, N. Farrington, and M. Missous, "Advanced Gunn diode as high power terahertz source for a millimetre wave high power multiplier," *Proc. SPIE*, vol. 7485, pp. 74 850I–74 850I–11, Sep. 2009.

[61] E. Alekseev, A. Eisenbach, D. Pavlidis, S. Hubbard, and W. Sutton, "Development of GaN-based Gunn-effect millimeter-wave sources," *Scientific Common*, 2010. [Online]. Available: http://www.scientificcommons.org/54287276.

[62] N. Alekseev, D. Malairov, and I. Bensen, "Generation of high-power oscillations with a magnetron in the centimeter band," *Proc. of the IRE*, vol. 32, no. 3, pp. 136–139, Mar. 1944.

[63] J. Osepchuk, "The magnetron and the microwave oven: A unique and lasting relationship," in *2010 Int. Conf. on the Origins and Evolution of the Cavity Magnetron (CAVMAG)*, Apr. 2010, pp. 46–51.

[64] Gil Wong Choi, Hae Jin Kim, Hyoung-Jong Kim, and Jin-Joo Choi, "The self-injection-locked magnetron," in *2008 IEEE International Vacuum Electronics Conference (IVEC 2008)*, Apr. 2008, pp. 445–446.

[65] T. Fleming, M. Lambrecht, and P. Mardahl, "Design and simulation of a mega-watt class nonrelativistic magnetron," *IEEE Trans. on Plasma Science*, vol. 40, no. 6, pp. 1563–1568, Jun. 2012.

[66] G. Caryotakis, "The klystron: A microwave source of surprising range and endurance," *Physics of Plasmas*, vol. 5, no. 5, pp. 1590–1598, Apr. 1998.

[67] A. M. Sessler and S. S. Yu, "Relativistic klystron two-beam accelerator," *Phys. Rev. Lett.*, vol. 58, pp. 2439–2442, Jun. 1987.

[68] K. Hendricks, P. Coleman, R. Lemke, M. Arman, and L. Bowers, "Extraction of 1 gw of RF power from an injection locked relativistic klystron oscillator," *Phys. Rev. Lett.*, vol. 76, pp. 154–157, Jan. 1996.

[69] C. Beard, "Review of available power sources," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 557, no. 1, pp. 276–279, Feb. 2006.

[70] J. Pierce, "Theory of the beam-type traveling-wave tube," *Proc. of the IRE*, vol. 35, no. 2, pp. 111–123, Feb. 1947.

[71] Young-Min Shin, L. Barnett, and N. Luhmann, "Phase-shifted traveling-wave-tube circuit for ultrawideband high-power submillimeter-wave generation," *IEEE Trans. on Electron Devices*, vol. 56, no. 5, pp. 706–712, Mar. 2009.

[72] S. Bhattacharjee, J. Booske, C. Kory, D. van der Weide, S. Limbach, S. Gallagher, J. Welter, M. R. Lopez, R. Gilgenbach, R. L. Ives, M. E. Read, R. Divan, and D. Mancini, "Folded waveguide traveling-wave tube sources for terahertz radiation," *IEEE Trans. on Plasma Science*, vol. 32, no. 3, pp. 1002–1014, Jun. 2004.

[73] W. Menninger, N. Robbins, D. Dibb, and D. Lewis, "Power flexible ka-band traveling wave tube amplifiers of up to 250-W RF for space communications," *IEEE Trans. on Elec-*

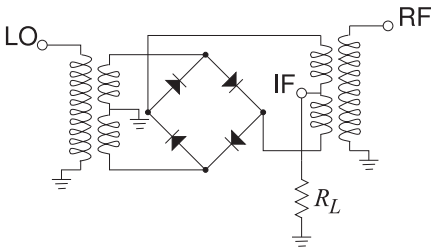
tron Devices, vol. 54, no. 2, pp. 181–187, Feb. 2007.

[74] J. Neilson, R. Ives, M. Caplan, M. Mizuhara, D. Marsden, and C. Kory, "High efficiency, terahertz, backward wave oscillators," in *The 29th IEEE International Conference on Plasma Science (ICOPS 2002)*, 2002, p. 171.

[75] R. Grow, J. Baird, K. Bunch, and R. C. Freudenberger, "Backward-wave oscillators for the frequency range from 300 GHz to 1 THz," in *2000 Int. Vacuum Electronics Conference,*, 2000.

6.14 Exercises

1. A mixer has an LO at 28.2 GHz. The mixer is used to convert a signal at 28.1 GHz to an IF at 100 MHz, and has a conversion loss of 13 dB and an image rejection of 40 dB. Two signals are presented to the mixer, one at 28.1 GHz with a power of 1 pW and the other at 28.3 GHz with a power of 10 μW. [Parallels Example 6.1]
   (a) What is the power of the (intended) signal at the IF in dBm?
   (b) What is the signal-to-interference ratio at the IF (ignoring noise)?
2. Consider the single-ended diode mixer in Figure 6-5(a).
   1. Develop a symbolic expression for the voltage at the test point. The diode is modeled by $i_D = a_1 v_D + a_2 v_D^2 + a_3 v_D^3$.
   2. What are the frequencies and amplitudes of the components of the spectrum at the test point?
3. A mixer in a receiver has a conversion loss of 16 dB. If the applied RF signal has an available power of 100 μW, what is the available power of the IF at the output of the mixer?
4. The RF signal applied to the input of a mixer has a power of 1 nW and the output of the mixer at the IF has a power level of 100 pW. What is the conversion loss of the mixer in decibels?
5. A mixer in a receiver has a conversion gain of 10 dB. If the applied RF signal has a power of 100 μW, what is the available power of the IF at the output of the mixer?
6. A mixer in a receiver has a conversion loss of 6 dB. If the applied RF signal has a power of 1 μW, what is the available power of the IF at the output of the mixer?
7. A mixer has an LO at 18 GHz. The mixer is used to convert a signal at 18.5 GHz to an IF at 500 MHz. Two signals are presented to the mixer, one at 18.5 GHz with a power of 100 nW and an interfering signal at 17.5 GHz with a power of 10 nW. If the image rejection is 20 dB and the conversion loss is 10 dB.
   (a) What is the signal power at the IF?
   (b) What is the interference power at the IF?
   (c) What is the signal-to-interference ratio (in decibels) at the IF?
8. A mixer has an LO at 100 GHz. The mixer is used to convert a signal at 110 GHz to an IF at 10 GHz. Two signals are presented to the mixer, one at 110 GHz with a power of 10 nW and an interfering signal at 90 GHz with a power of 5 nW. If the image rejection is 40 dB and the conversion loss is 20 dB, what is the signal-to-interference ratio (in decibels) at the IF?
9. A mixer has an LO at 18 GHz. The mixer is used to convert a signal at 18.5 GHz to an IF at 500 MHz. The RF signal at 18.5 GHz has a power of 100 pW. In addition, noise with a power of 1 pW is applied to the mixer at 18.5 GHz and 17.5 GHz (1 pW at 18.5 GHz and 1 pW at 17.5 GHz). If the image rejection is 6 dB and the conversion loss is 10 dB. Ignore noise contributions from the mixer. What is the SNR (in decibels) at the IF?
10. A mixer in a communication system has an LO at 5.5 GHz. The mixer is used to convert a 10 MHz bandwidth signal at 5.6 GHz to an IF at 100 MHz. The RF signal at 5.6 GHz has a power of 100 pW. The image rejection is ideal and the conversion loss is 10 dB. The mixer has a single-sideband noise figure of 6 dB.
    (a) What is the noise power at the input if the source is held at standard temperature (290 K)?
    (b) What is the input SNR (in decibels)?
    (c) What is the SNR (in decibels) at the IF?
11. The double-balanced ring diode mixer shown below has the special characteristic that the LO and RF tones are suppressed at the IF output port. Develop a symbolic expression for the voltage at the IF port. The diodes are matched and are modeled by $i_D = a_1 v_D + a_2 v_D^2 + a_3 v_D^3$. The LO voltage, at the LO terminal, is $v_{\mathrm{LO}} = A\cos(\omega_{\mathrm{LO}} t)$ and the RF voltage, at the RF terminal, is $v_{\mathrm{RF}} = B\cos(\omega_{\mathrm{RF}} t)$. Consider a 1:1 winding ratio. That is, the number of windings on the secondary on each side of the center tap is equal to the number of windings on the primary.



12. A diode double-balanced mixer has an LO at 100 GHz and has an input RF signal of 101 GHz. What will be the frequencies of the main signals at the IF?
13. The phase noise of an oscillator was measured as -120 dBc/Hz at 10 kHz offset. What is the normalized phase noise at 1 MHz offset, assuming that the phase noise power varies as the inverse of frequency?
14. The phase noise of an oscillator was measured as -130 dBc/Hz at 10 kHz offset. What is the normalized phase noise at 1 MHz offset, assuming that the phase noise power varies as the inverse of frequency?
15. The phase noise of an oscillator was measured as -125 dBc/Hz at 100 kHz offset. What is the normalized phase noise at 1 MHz offset, assuming that the phase noise power varies as the square of the inverse of frequency?
16. The phase noise of an oscillator was measured as -125 dBc/Hz at 100 kHz offset. What is the normalized phase noise at 1 MHz offset, assuming that the phase noise power varies inversely with frequency offset?
17. When 0 V is applied to a VCO, the output frequency is 1 GHz. When the input to the VCO is 10 mV, the sinusoidal output of the VCO has a frequency of 1.01 GHz. What is the tuning gain of the VCO?
18. If a sinusoidal voltage is applied to the input of an analog VCO, describe the signal at the output of the VCO.
19. Describe the design of a times-two frequency divider using a frequency multiplier based on a diode and one or more bandpass filters. That is, sketch the circuit at the block diagram level.
20. Describe the design of a times-three frequency divider using a frequency multiplier based on a diode and one or more bandpass filters. That is, sketch the circuit at the block diagram level.

6.14.1 Exercises By Section

†challenging, ‡very challenging

§6.2 $1, 2^{\dagger}, 3, 4, 5, 6, 7^{\dagger}, 8^{\dagger}, 9^{\dagger}, 10^{\dagger}$
§6.3 $11^{\ddagger}, 12$
§6.4 $13^{\dagger}, 14^{\dagger}, 15^{\dagger}, 16$
§6.5 $17, 18$
§6.9 $19, 20$

6.14.2 Answers to Selected Exercises

3 -26 dBm
5 0 dBm
7(b) 10 pW
9 19.0 dB
10(b) 28 dB
12 One signal at 1 GHz
17 1 GHz/V

CHAPTER 7

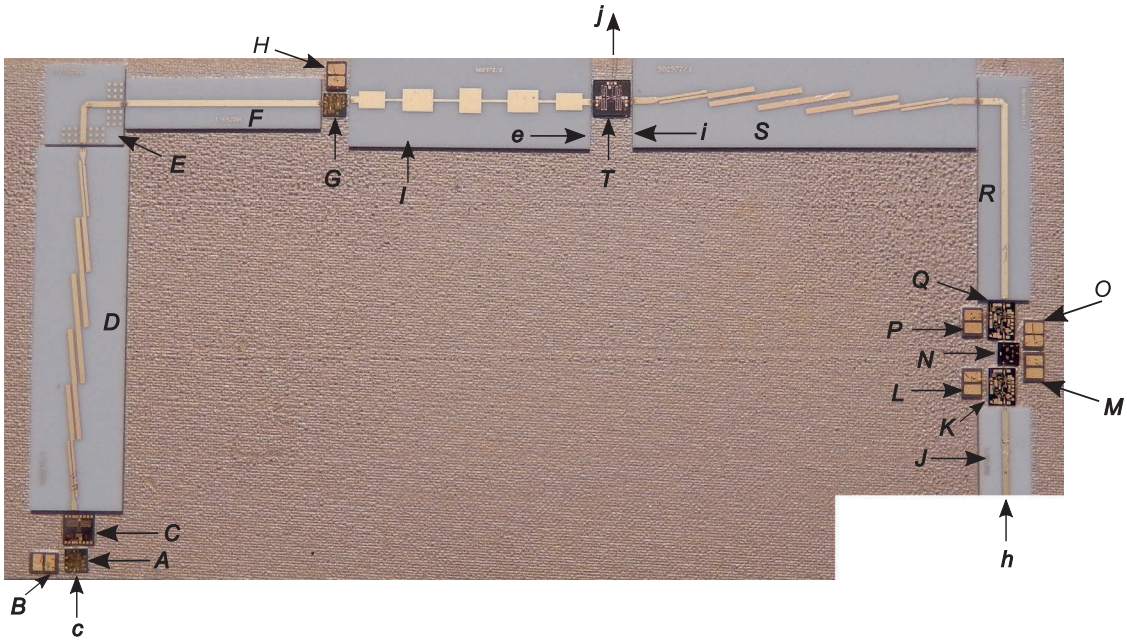
Cascade of Modules

7.1 Introduction 223
7.2 Nonlinear Distortion of a Cascaded System 225
7.3 Cascaded Module Design Using the Budget Method 231
7.4 Cascaded Module Design Using the Contribution Method 232
7.5 Case Study: High Dynamic Range Down-Converter Design ... 236
7.6 Case Study: Analysis of a 15 GHz Receiver 238
7.7 Case Study: Frequency Planning of a Transceiver 241
7.8 Summary 247
7.9 References 248
7.10 Exercises 248

7.1 Introduction

Design of a receiver or transmitter circuit requires the design of a cascade of modules that achieves optimum dynamic range while managing DC power consumption. As an example the frequency conversion stage of a receiver is shown in Figure 7-1. The microstrip filters and transmission lines are fabricated on alumina substrates. The semiconductor dies and the chip decoupling capacitors as well as the alumina modules are epoxied to a mat that was screen-printed on a brass housing. The mat provides a stress-relieving (allowing for differences in thermal coefficients of expansion) and conducting interface between the alumina and semiconductor substrates and the brass housing. The modules and dies are interconnected by bond wires arranged as two or more bond wires in parallel to reduce inductance. The bond wires attach to the pads on the decoupling capacitors. These modules, with the exception of filters, are described in this chapter.

The chapter begins with investigation of a cascade of modules given the distortion characteristics of the individual modules. Coupled with an earlier treatment of noise of a cascaded system this defines dynamic range. A microwave systems comprising a cascade of mostly two-port modules and must be designed to simultaneously minimize noise, distortion, DC power consumption, spurious emissions, and maximize dynamic range. There is necessarily a trade-off of these performance parameters and this trade-off is at the heart of system design. What makes this particularly challenging is that the two-port networks are designed separately and there



Modules		Modules	
A	Driver amplifier	***N***	Variable attenuator
B,H	Bias decoupling capacitor	***Q***	Low noise amplifier
C	×2 frequency multiplier	***R***	Transmission line
D	Edge-coupled bandpass filter	***S***	Edge-coupled bandpass filter
E	Mitered bend	***T***	Subharmonic mixer
F	Transmission line		
G	×2 frequency multiplier	Signal Path	
I	Lowpass filter	**e**	Frequency multiplied LO
J	Transmission line	**h**	RF input
K	Low noise amplifier	**i**	Reference LO input
L,M,O,P	Bias decoupling capacitors	**j**	IF output

Figure 7-1: The frequency conversion portion of the 15 GHz receiver shown in Figure 7-12. The RF input at **h** is centered at 15 GHz and is amplified and filtered before being presented to athe subharmonic mixer at **T**. The reference LO at **c** is tunable from 1601 to 1742 MHz and is frequency multiplied twice, at **C**and **G**, and then presented as the LO to the harmonic mixer at at **T**. The LO ranges from 6403 to 6968 MHz. The IF output at **j** ranges from 465 to 1595 MHz.

is incomplete characterization of performance. Indeed the only accurate way of determining overall system performance would be to have the circuit-level schematics of each module and then perform a whole system time-domain circuit-level siulation using the actual modulated signals. Even if all the details were available the simulation time would be prohibitive and a time-domain simulation is unlikely to have the necessary fidelity. This is true even if a large part of a system is implemented on the same chip. Thus system design must use a methodical process using incomplete module-level parameters such as the 1 dB gain compression, noise figure, and third-order intercept characterizations. System design is often conservative designing for worst case. Sometimes this is too conservative and allowances must be made.

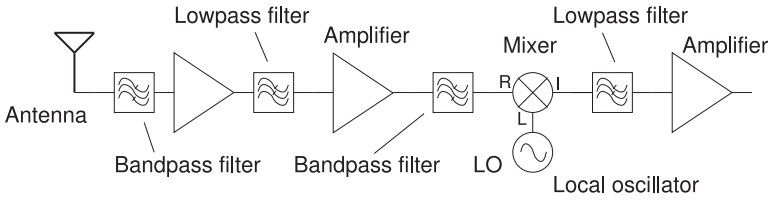


Figure 7-2: Receiver as a cascade of modules.

Testing and adjustment of fabricated prototypes is essential.

This section describes two approaches to cascaded module design; the budget method and the contribution method. In the budget method gain and noise performance metrics are initially assigned to each stage in a cascade and then these are adjusted in a system simulator to iterate towards an optimum solution. In the contribution method the focus is on dynamic range and each critical module is assigned the same dynamic range and from this an initial assignment of noise figure, gain, and required linearity is made for each module. The optimization of system performance proceeds as with the budget method. Generally this optimization must be performed manually using experience as often it is difficult to quantify what an optimum is. Also some design decisions require the acquisition of higher cost modules and if a module is low cost then it makes sense to assign high dynamic range to such a module to reduce the requirements on other higher cost modules. Perhaps a module must be redesigned or designed in-house instead of being externally sourced. These are not simple decisions. In the end, the system with cascaded modules must be cost and performance competitive, and reducing time to market is a competitive advantage.

7.2 Nonlinear Distortion in a Cascaded System

This section builds on the distortion analysis of two-port networks in Section 4.5 and addresses the determination of the nonlinear metrics of a cascaded system such the receiver shown in Figure 7-2. The filters and interstage matching networks eliminate harmonics from the system but, unfortunately, allow in-band and close-in out-of-band distortion components to pass through the system. The main metrics that describe nonlinear performance are the power levels at the 1 dB gain compression point and at the third-order intercept (IP3) point. In some systems, such as direct-conversion receivers, the second-order intercept point (IP2) is also important. These metrics relate to discrete tones and the correlated distortion generated in the different stages. This contrasts with the calculation of noise of a cascaded system where the noise added by stages is uncorrelated.

7.2.1 Gain Compression in a Cascaded System

When two amplifier stages are cascaded it is necessary to determine the 1 dB gain compression point of the cascade. Consider the cascade in Figure 7-3. The two stages have linear power gains G_1 and G_2, and 1 dB compression points $P_{1,1\text{dB}}$ and $P_{2,1\text{dB}}$, respectively. The total linear power gain of the system is $G^{\text{T}} = G_1 \cdot G_2$. If the Taylor series expansion of the input-output characteristics of the first stage in the cascade is

$$v_{1o}(t) = a_o + a_1 v_{1i}(t) + a_2 v_{1i}^2(t) + a_3 v_{1i}^3(t) + \dots , \tag{7.1}$$

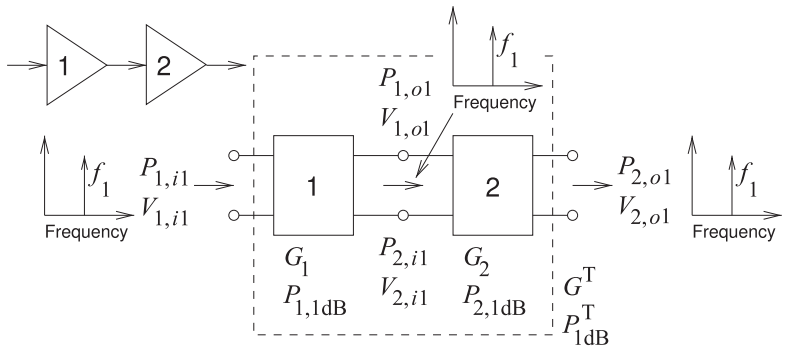


Figure 7-3: Cascade of two stages used in determining the total 1 dB compression point of the system.

then at the interface of the two stages the amplitude of the tone at f_1 due to a single-tone input, $v_i(t) = V_i \cos(\omega_1 t)$, is $v_{1o}(t) = V_{1o} \cos(\omega_1 t)$, where

$$V_{1o} = a_1 V_i + \tfrac{3}{4} a_3 V_i^3. \tag{7.2}$$

The harmonics are ignored, as design of the interstage matching network would preferably be lowpass to eliminate the passage of harmonics. The input of the second stage is $v_{2i}(t) = v_{1o}(t)$, with the Taylor series expansion of the second stage being

$$v_{2o}(t) = b_o + b_1 v_{2i}(t) + b_2 v_{2i}^2(t) + b_3 v_{1i}^3(t) + \dots . \tag{7.3}$$

Thus the component of the final output at f_1 will be

$$V_o = a_1 b_1 V_i + \tfrac{3}{4} b_1 a_3 V_i^3 + \tfrac{3}{4} a_1 b_3 V_i^3 + \dots . \tag{7.4}$$

That is, the input signal is multiplied by the overall linear gain of the amplifier; this is the first term on the right in Equation (7.4), and the last two terms result in gain compression. If the stages are identical (i.e., $a_1 = b_1$ and $a_3 = b_3$) the two gain compression terms will be identical. Thus the two stages will contribute equally to gain compression.

Note that in Equation (7.4) voltages add to yield the overall gain compression. In general the two stages will not be identical, they will have different gain and 1 dB compression points, however, this observation (i.e., that the voltages add) does enable an approximate expression to be developed for the 1 dB compression point of a cascaded system.

Examination of Equation (7.4) leads to a general formula for the total 1 dB compression point, $P^{\text{T}}_{1\text{dB}}$, of the two-stage cascade in Figure 7-3:

$$\left(P^{\text{T}}_{1\text{dB}}\right)^{-\frac{1}{2}} \approx \left(G_2 P_{1,1\text{dB}}\right)^{-\frac{1}{2}} + \left(P_{2,1\text{dB}}\right)^{-\frac{1}{2}} . \tag{7.5}$$

Note that here the gains and the powers are absolute quantities and not in decibels. For example, $P^{\text{T}}_{1\text{dB}}$ is the power in watts at the 1 dB gain compression point.

Equation (7.5) is a very conservative estimate of the 1 dB gain compression point of the amplifier and flags the lowest power levels at which gain compression of the cascade could occur. The situation could be better depending on the phasing of the distortion components but conservatism is very important in system design. The worst case is what matters in specifying system performance. Getting the effect of phasing right in system design requires a circuit-level simulation but when working with modules

such information is rarely available. Equation (7.5) cannot be directly derived from the voltage expansion as the power gain of microwave amplifiers is due to difference of input and output impedance levels in addition to voltage gain. It should be noted that a microwave amplifier deviates from linearity following a sharper, tanh-like, response than that described by a low-order polynomial model of linearity. Thus simply cascading low-order polynomials describing each stage is not a viable option.

If required and if a more realistic estimate is required and can be justified, then a less conservative formula that can be used to estimate the gain compression level of two cascaded stages is

$$\left(P_{1\text{dB}}^{\text{T}}\right)^{-1} \approx \left(G_2 P_{1,1\text{dB}}\right)^{-1} + \left(P_{2,1\text{dB}}\right)^{-1}. \tag{7.6}$$

It is usually better to use the more conservative estimate of compression in Equation (7.5).

EXAMPLE 7.1 Gain Compression of a Two-Stage Amplifier

The first stage of a two-stage amplifier in a transmitter has a gain $G_1 = 20$ dB and an output 1 dB gain compression power $P_{1,1\text{dB}} = 10$ dBm. The second stage has a gain $G_2 = 6$ dB and an output 1 dB gain compression power $P_{2,1\text{dB}} = 20$ dBm.

(a) What is the linear gain of the two-stage amplifier?
(b) What is the gain of the two-stage amplifier at the 1 dB gain compression power?
(c) What is the 1 dB gain compression power of the cascaded system?

Solution:

(a) When the gain of an amplifier stage is given without qualification it should be assumed to be the linear gain, that is, the gain at small signal levels. So the total linear power gain of the two-stage amplifier is

$$G^{\text{T}} = G_1 G_2 = 20\text{ dB} + 6\text{ dB} = 26\text{ dB}.$$

(b) The compressed gain will be 1 dB less, that is, $G^{\text{T}} = 25$ dB.
(c) The overall gain compression power, $P_{1\text{dB}}^{\text{T}}$ is, approximately obtained using Equation (7.5):

$$\left(P_{1\text{dB}}^{\text{T}}\right)^{-\frac{1}{2}} = \left(G_2 P_{1,1\text{dB}}\right)^{-\frac{1}{2}} + \left(P_{2,1\text{dB}}\right)^{-\frac{1}{2}} = \left(10^{(6/10)} 10^{(10/10)}\right)^{-\frac{1}{2}} + \left(10^{(20/10)}\right)^{-\frac{1}{2}}$$

$$P_{1\text{dB}}^{\text{T}} = \left[(3.981 \cdot 10)^{-\frac{1}{2}} + (100)^{-\frac{1}{2}}\right]^{-2} \text{ mW} = 14.97\text{ mW} = 11.8\text{ dBm}. \tag{7.7}$$

A quick check is as follows.
If Stage 1 dominates gain compression, the output power at the 1 dB compression is $P_{1,1\text{ dB}} = 10$ dBm multiplied by the linear power gain of the second stage, that is, $P_{1\text{dB}}^{\text{T}} = G_2 P_{1,1\text{ dB}} = 6\text{ dB} + 10\text{ dBm} = 16\text{ dBm}$.
If Stage 2 dominates compression, the output power at the 1 dB compression of the two-stage amplifier is just that of Stage 2: $P_{1\text{dB}}^{\text{T}} = 20$ dBm.

EXAMPLE 7.2 Second Example of Gain Compression of a Two-Stage Amplifier

The first stage of a two-stage amplifier in a transmitter has a gain $G_1 = 13$ dB and an output 1 dB gain compression power $P_{1,1\text{dB}} = 10$ dBm. The second stage gain is $G_2 = 10$ dB and the output 1 dB gain compression power is $P_{2,1\text{dB}} = 20$ dBm. What is the 1 dB gain compression power of the two-stage amplifier system?

Solution:

The total gain compression power, $P_{1\text{dB}}^{\text{T}}$, is approximately obtained using Equation (7.5):

$$\left(P_{1\text{dB}}^{\text{T}}\right)^{-\frac{1}{2}} = (G_2 P_{1,1\text{dB}})^{-\frac{1}{2}} + (P_{2,1\text{dB}})^{-\frac{1}{2}} = \left(10^{\frac{10}{10}} 10^{\frac{10}{10}}\right)^{-\frac{1}{2}} + \left(10^{\frac{20}{10}}\right)^{-\frac{1}{2}} \text{ (mW)}^{-\frac{1}{2}}$$

$$P_{1\text{dB}}^{\text{T}} = \left[(100)^{-\frac{1}{2}} + (100)^{-\frac{1}{2}}\right]^{-2} \text{ mW} = 25.00 \text{ mW} = 13.98 \text{ dBm}. \tag{7.8}$$

A quick check is as follows.
If Stage 1 dominates compression, the output power at 1 dB compression of the two-stage amplifier is $P_{1,1\text{dB}}$ multiplied by the linear power gain of the second stage, taht is $P_{1\text{dB}}^{\text{T}} = G_2 P_{1,1\text{ dB}} = 10 \text{ dBm} + 10 \text{ dB} = 20 \text{ dBm}$.
If Stage 2 dominates compression, $P_{1\text{dB}}^{\text{T}} = P_{2,1\text{dB}} = 20$ dBm. Thus neither stage dominates compression.

A final note is that the calculation of the 1 dB compression point when both stages in a two-stage system contribute equally to gain compression is approximate as the actual compression characteristic is complex and a third-order Taylor series does not capture the total nonlinear response [1, 2].

The response of a multistage system can be extrapolated from the treatment here for two stages. For an m-stage cascade,

$$\left(P_{1\text{dB}}^{\text{T}}\right)^{-\frac{1}{2}} = (G_m \ldots G_2 P_{1,1\text{dB}})^{-\frac{1}{2}} + \ldots + \left(G_2 P_{(m-1),1\text{dB}}\right)^{-\frac{1}{2}} + (P_{m,1\text{dB}})^{-\frac{1}{2}}. \tag{7.9}$$

7.2.2 Intermodulation Distortion in a Cascaded System

The two-stage system shown in Figure 7-4 will be used here to determine the total intermodulation response as described by the third-order input and output intercept points, IIP3 and OIP3, respectively. The development is based on the analysis of intermodulation distortion in Section 4.5.3 and is called the cascade intercept method. One version of the method is called the organized cascade intercept method, in which the worst-case situation

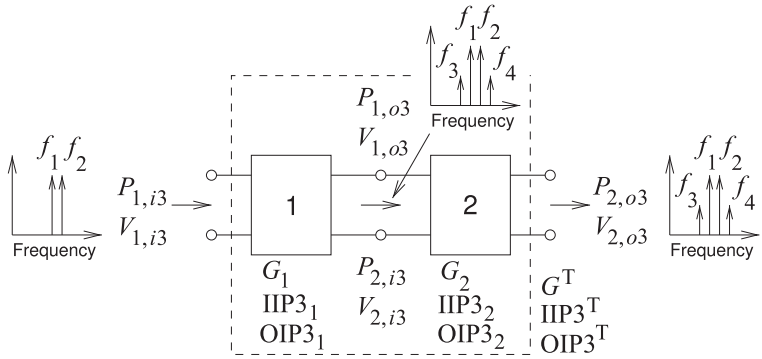


Figure 7-4: Cascade of two stages used in determining the third-order intercept point of a cascaded system.

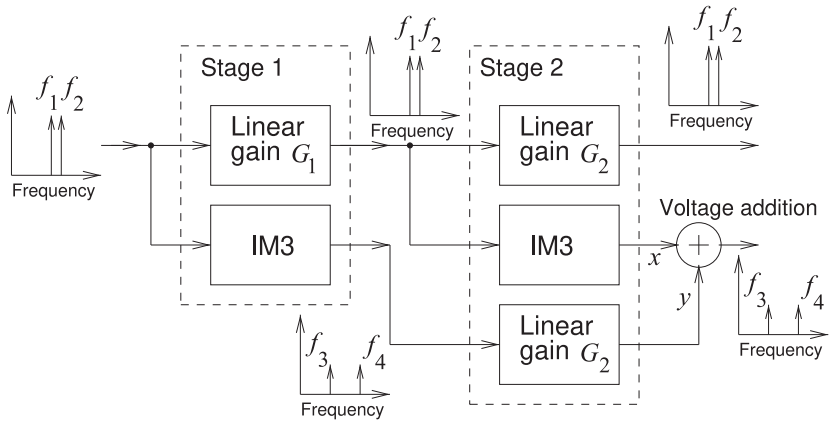


Figure 7-5: Signal flow for calculating the intermodulation distortion in the cascade of two stages.

is assumed, in that the IM3 produces from each stage in the cascade add in phase. The other method is the unorganized cascade intercept method in which the phases of IM3 produce by each stage are unknown, and assumes that the most likely overall IM3 distortion will be obtained by assuming a random phase relationship of the IM3 contributions.

The third-order intercept point, IP3, is determined by extrapolating the small signal gain and IM3 responses. Since IM3 is small, the signal• flow in the two-stage system is as shown in Figure 7-5. One path in the flow linearly amplifies the input two-tone signal in Stages 1 and 2. The third-order nonlinearity of Stage 1 creates low-level IM3 that is linearly amplified by the second stage. A second set of IM3 signals is produced by the IM3 of the second stage operating on the two-tone signal amplified by Stage 1. The voltages of the two IM3 signals add constructively. A complication is that the phases of the two IM3 signals, at x and y, are not known. This leads to two approaches to estimating of the overall distortion.

Organized Cascade Intercept Method

The worst-case situation is that the IM3 voltages at x and y are in phase so that the voltages add. Then the total IM3 distortion is obtained by adding the IM3 voltages generated from each stage. Determining total OIP3, $\mathrm{OIP3^T}$ is the inverse problem. This cannot be solved precisely, but a good estimate for $\mathrm{OIP3^T}$ is obtained from [3, 4]

$$\left(\mathrm{OIP3^T}\right)^{-\frac{1}{2}} \approx \left(G_2\mathrm{OIP3}_1\right)^{-\frac{1}{2}} + \left(\mathrm{OIP3}_2\right)^{-\frac{1}{2}}. \tag{7.10}$$

Note that OIP3 is a power and G is a power gain. Now the total gain of the cascade $G^{\mathrm{T}} = G_1 G_2$, and since $\mathrm{OIP3^T} = G^{\mathrm{T}}\mathrm{IIP3^T}$, where $\mathrm{OIP3^T}$, the input intercept of the cascade, can be written

$$\left(\mathrm{IIP3^T}\right)^{-\frac{1}{2}} \approx \left(\mathrm{IIP3}_1\right)^{-\frac{1}{2}} + \left(\mathrm{IIP3}_2/G_1\right)^{-\frac{1}{2}}. \tag{7.11}$$

Generalizing these results, for an m-stage cascade the overall OIP3 is

$$\left(\mathrm{OIP3^T}\right)^{-\frac{1}{2}} \approx \left(G_m \ldots G_2\mathrm{OIP3}_1\right)^{-\frac{1}{2}} + \ldots \left(G_m\mathrm{OIP3}_{(m-1)}\right)^{-\frac{1}{2}} + \left(\mathrm{OIP3}_m\right)^{-\frac{1}{2}}. \tag{7.12}$$

Since $\mathrm{IIP3} = \mathrm{OIP3}/G$, the general result for multiple cascaded stages can be written

$$\left(\frac{1}{\mathrm{IIP3^T}}\right)^{\frac{1}{2}} \approx \left(\frac{1}{\mathrm{IIP3}_1}\right)^{\frac{1}{2}} + \ldots \left(\frac{G_{(m-2)} \ldots G_1}{\mathrm{IIP3}_{(m-1)}}\right)^{\frac{1}{2}} + \left(\frac{G_{(m-1)} \ldots G_1}{\mathrm{IIP3}_m}\right)^{\frac{1}{2}}. \tag{7.13}$$

The organized cascade intercept method often provides an overly conservative (i.e., far too low) estimate of $\mathrm{OIP3^T}$ and $\mathrm{IIP3^T}$ distortion.

Unorganized Cascade Intercept Method

If the phases of the stages are random (or perhaps unknown), then it is reasonable to add the powers of the distortion terms. This is an approximation as the IM3 signals are still correlated, but this approach has been found to provide a useful measure in design, then

$$\left(\mathrm{OIP3^T}\right)^{-1} \approx \left(G_2 \mathrm{OIP3}_1\right)^{-1} + \left(\mathrm{OIP3}_2\right)^{-1}, \tag{7.14}$$

and again OIP3 is a power and G is a power gain. For an m-stage cascade with random IM3 phase the total OIP3, $\mathrm{OIP3^T}$, is obtained from

$$\left(\mathrm{OIP3^T}\right)^{-1} \approx \left(G_m \ldots G_2 \mathrm{OIP3}_1\right)^{-1} + \ldots + \left(G_2 \mathrm{OIP3}_{(m-1)}\right)^{-1} + \left(\mathrm{OIP3}_m\right)^{-1} \tag{7.15}$$

and the total IIP3, $\mathrm{IIP3^T}$, is obtained from (since $\mathrm{IIP3} = \mathrm{OIP3}/G$)

$$\left(\frac{1}{\mathrm{IIP3^T}}\right) \approx \left(\frac{1}{\mathrm{IIP3}_1}\right) + \left(\frac{G_1}{\mathrm{IIP3}_2}\right) + \cdots + \left(\frac{G_{(m-1)} \ldots G_1}{\mathrm{IIP3}_m}\right). \tag{7.16}$$

Summary

The cascade intercept method has led to two sets of results for IM3 distortion. The first, from the organized cascade intercept method, is the worst-case situation in which the IM3 distortion of each stage combines in the worst possible way. This yielded the overall IIP3 and OIP3 results of Equations (7.12) and (7.13). The second, from the unorganized cascade intercept method, assumes that the phases of the IM3 from each stage are randomly related, perhaps the best estimate that can be made without a circuit simulation. This yielded the overall IIP3 and OIP3 results of Equations (7.15) and (7.16).

It is interesting to speculate what would happen if the stages were designed so that the IM3 contributions at x and y (referring to Figure 7-5) of the first and second stages were $180°$ out of phase but with the same magnitude. If this could be done the IM3 contributions at the output would be canceled. Indeed, it is possible to do this to a limited extent. The phase of the IM3 signals depends on the signal level and so changes over the signal range. Careful design, and only when there is complete control over the design and integration of the stages, enables the IMD contributions to be partially canceled over a range of signal levels, as shown in Figure 7-6. This extends the dynamic range of the cascaded system, and the simple OIP3 and IIP3 metrics are not sufficient to capture this complexity.

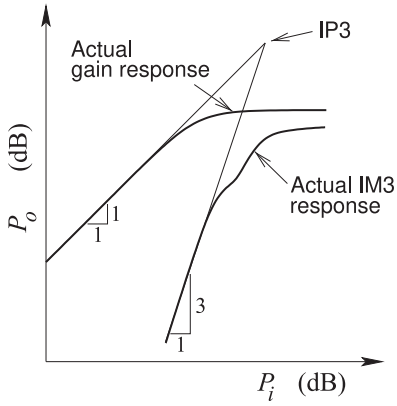


Figure 7-6: The gain and IM3 responses of two cascaded stages showing the partial cancellation of the IM3 response over a range of input signal levels.

EXAMPLE 7.3 Intermodulation Distortion of a Two-Stage Amplifier

The first stage of a two-stage amplifier in a transmitter has a gain of $G_1 = 20$ dB and an output third-order intercept point, $\mathrm{OIP3}_1$, of 30 dBm. The second stage has a gain of $G_2 = 20$ dB and an OIP3 of 40 dBm. Assume that the IMD contributions of each stage are in phase. What is the OIP3 of the cascaded system?

Solution:

Using the organized cascade intercept method, the total OIP3, $\mathrm{OIP3}^{\mathrm{T}}$, in milliwatts is obtained using Equation (7.10)

$$\left(\mathrm{OIP3}^{\mathrm{T}}\right)^{-\frac{1}{2}} \approx \left(10^{(20/10)}10^{(30/10)}\right)^{-\frac{1}{2}} + \left(10^{(40/10)}\right)^{-\frac{1}{2}} = \left(10^5\right)^{-\frac{1}{2}} + \left(10^4\right)^{-\frac{1}{2}}. \tag{7.17}$$

Thus $\mathrm{OIP3}^{\mathrm{T}} = 5772\ \mathrm{mW} = 37.6\ \mathrm{dBm}.$ (7.18)

This is the worst-case situation. It is worth comparing this to a calculation where the phases of the IMD contributions are unknown. Then, using Equation (7.14),

$$\left(\mathrm{OIP3}^{\mathrm{T}}\right)^{-1} \approx \left(10^{(20/10)}10^{(30/10)}\right)^{-1} + \left(10^{(40/10)}\right)^{-1} = 10^5 + 10^4. \tag{7.19}$$

Thus $\mathrm{OIP3}^{\mathrm{T}} = 9091\ \mathrm{mW} = 39.6\ \mathrm{dBm}.$ (7.20)

7.3 Cascaded Module Design Using the Budget Method

The budget method is used to design RF cascade systems, particularly the design of receiver and transmitter cascaded systems, for specified gain, noise, and distortion. In the budget method, initial assignments of gain and noise performance are made to each stage in a cascade. This approach is based on the calculation of the total noise figure and gain of a system from the parameters of individual subsystem stages (see Section 4.3.1). This is coupled with the calculation of nonlinear distortion in cascaded stages (in Section 7.2), to calculate the spurious free dynamic range (SFDR) and make assignments for noise and distortion contributions of each stage [5]. The SFDR is related to the gain compression, noise, and nonlinear distortion metrics in Figure 7-7. The system optimization goal is, in general, to choose and adjust modules to maximize overall SFDR subject to the constraints of cost and availability of modules with the required attributes. For example, the gain of most amplifier modules can be adjusted and so change the

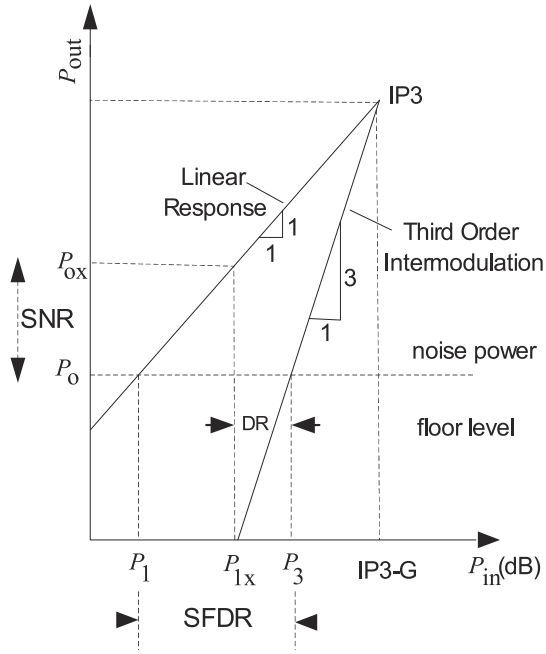


Figure 7-7: Output power versus input power of a stage or system. Extrapolations of the 1:1 linear response and the 3:1 third-order intermodulation response intersect at the IP3 point. DR = dynamic range, SFDR = spurious free dynamic range, SNR = required minimum signal-to-noise ratio.

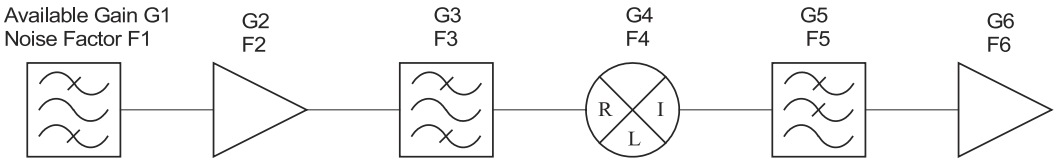


Figure 7-8: Cascade system for receive down converter or transmit up converter.

nonlinear distortion performance of amplifier stages. In the budget method the noise and nonlinear distortion metrics of each stage are set based on experience.

Considering the stages in the cascade shown in Figure 7-8, generally maximum gain is assigned to the first active stage, and so the first stage contributes the most to the system noise figure. However, as a consequence active devices in latter stages nearly always require significantly higher linearity performance than may be necessary to meet the system SFDR objective. Still the budget method is a systematic way to design a cascaded system. Experience can also be used in the budget assignments. For example, in transmitter cascade design the intent might be to provide as high an intercept as possible in the lower-frequency up-conversion stages. This approach is based on devices operating at lower frequencies having very good linearity and therefore being most cost effective in achieving high IP3.

7.4 Cascaded Module Design Using the Contribution Method

The contribution method is an approach to the design of cascade RF systems for maximum SFDR [6] rather than separate treatment of noise and nonlinear

distortion. The contribution method provides a good initial assignment of the noise figure, gain, and required linearity to individual stages and enables informed assessment of trade-offs during system design.

In the budget method, noise performance is largely determined by the first stage and the nonlinear performance by latter stages. This often results in suboptimum system design. One of the costs is often the use of more expensive stages than necessary or higher linearity requirements of some stages leading to increased supply power.

The contribution method assigns to each active stage a percentage contribution to the overall SFDR. Contributions of the stages are balanced as no single stage dominates noise or distortion, and all contributions to SFDR are equally weighted. In realizing a cascaded system, typically one or more offending stages either dominate the SNR or significantly degrade overall distortion. In some cases a redistribution of stage parameters (such as gain) reduces the degradation. In other cases a change in architecture is required.

The mathematics behind the contribution approach is based on an extension of the noise and distortion analyses of cascaded stages. The noise performance of a stage is characterized by its noise factor $F = N_o/(GN_i)$, where G is the available power gain of the stage, N_o is the output noise power, and N_i is the input noise power of a resistor held at standard temperature (290 K). Distortion is characterized by the amount of intermodulation distortion (IMD) produced in a two-tone test and specifically the third-order intercept point (IP3) (see Figure 7-7). In the following development a passive linear stage such as a filter or attenuator is combined with the following active stage.

7.4.1 Noise Contribution

Friis's formula yields the noise factor of a cascaded system and was derived in Section 4.3.1. This section presents an alternative development of Friis's formula with interim steps that can be used in calculating the dynamic range contributions of each stage. The analysis here was originally presented in [6].

First consider the contribution of individual stages to system noise, and assume that all stages are matched. The nth stage, with all ports having equal bandwidth B, has output noise power (from Equation (4.24)

$$N_{no} = F_n G_n k T_0 B, \tag{7.21}$$

where k is the Boltzmann constant, and F_n and G_n are the noise factor and available power gain of the nth stage, respectively. The excess noise power of successive stages is additional to that of the first. So, while the output noise power contribution of the first stage is

$$N_{1o} = F_1 G_1 k T_0 B, \tag{7.22}$$

the excess output noise power of the second stage is

$$N_{2o} = (F_2 - 1) k T_0 B G_2. \tag{7.23}$$

Note that the output noise of the first stage includes the noise contribution of the source resistance held at the standard temperature. However, this is not included in the noise contributed by the second or latter stages. The total noise power at the output of a two-stage cascade is

$$N_{2o}^{\mathrm{T}} = (F_2 - 1) k T_0 B G_2 + N_{1o} G_2. \tag{7.24}$$

Then the total noise power at the output of the mth stage is

$$N_{mo}^{\mathrm{T}} = \sum_{n=2}^{m} \left[(F_n - 1) k T_0 B \prod_{i=2}^{n} G_i \right] + F_1 k T_0 B \prod_{n=1}^{m} G_n. \tag{7.25}$$

Thus an m-stage cascade has total cascaded system noise factor $F^{\mathrm{T}} = N_{mo}^{\mathrm{T}} / \left(G^{\mathrm{T}} N_{1i} \right)$, with G^{T} being the total cascaded available gain and N_{1i} the noise power input to the first stage. In terms of the parameters of individual stages

$$F^{\mathrm{T}} = F_1 + \sum_{n=2}^{m} \frac{F_n - 1}{\prod_{i=n}^{m} G_{i-1}}. \tag{7.26}$$

The link between Equations (7.25) and (7.26) enables the noise contribution of each stage to be determined. Treated separately, Equation (7.25) provides the output noise power of a cascade such as a receiver. Although the extension of Equation (7.25) to Equation (7.26) is normally associated with the derivation of the receiver noise factor, it can also be used in transmitter noise analysis.

Now the noise contribution of a stage can be defined. The gain accumulated at the jth stage (the total cascade gain up to and including the jth stage) is

$$G_j^{\mathrm{A}} = \prod_{n=1}^{j} G_n \tag{7.27}$$

(and so $G^{\mathrm{T}} = G_m^{\mathrm{A}}$ for an m-stage cascade). The fractional noise contribution of the jth stage to the total output noise is then defined as

$$C_j^N = \text{ stage noise contribution} \equiv \left[\frac{F_j - 1}{\left(G_j^{\mathrm{A}} / G_j \right) F^{\mathrm{T}}} \right]. \tag{7.28}$$

This is one component of the SFDR contribution of a stage.

7.4.2 Intermodulation Contribution

The intermodulation distortion of a cascade is assessed using the cascade intercept method (see Section 7.2.2). The usual design approach of establishing the SNR in the early stages of the cascade (this is the budget method) results in stages further along the cascade needing to have higher intercept values. Consequently these stages will generally have additional power consumption, as this is usually required to increase IIIP3 (or OIP3). Trade-offs of the contributions of individual stages to distortion and noise will lead to constrained total power consumption while still achieving the required system SFDR.

For receiver systems the contribution to nonlinear distortion is captured by IIP3 and for all stages this will be referred to the input of the cascaded system. Keeping with the nomenclature of the previous section, the accumulated IP3 of the jth stage referred to the system input is

$$\mathrm{IIP3}_j^{\mathrm{A}} = \mathrm{IIP3}_j / G_{j-1}^{\mathrm{A}} \tag{7.29}$$

and is $\text{IIP3}^{\text{A}}_{\text{dBm},j}$ when expressed in dBm. Note that all prior gain and loss up to the jth stage modifies the intercept when it is referred to the input of the cascaded stages. Combining the contribution of individual stages using Equation (7.16) yields the total system IIP3:

$$\text{IIP3}^{\text{T}} = \left(\sum_{n=1}^{m} \frac{1}{\text{IIP3}^{\text{A}}_{n}} \right)^{-1}. \tag{7.30}$$

This leads to the definition of the fractional contribution of the jth stage to the system IIP3:

$$C_j^{\text{IIP3}} = \text{IIP3 contribution} \equiv \frac{\text{IIP3}^{\text{T}}}{\text{IIP3}^{\text{A}}_{j}}. \tag{7.31}$$

This is the second and final component of the SFDR contribution of a stage.

7.4.3 Design Methodology for Maximizing Dynamic Range

In this section a design methodology is developed for trading off the performance of each stage in maximizing the SFDR of the RF down-converter module shown in Figure 7-8. A central component of the methodology is the use of a stage contribution graph that indicates performance of each stage. This graph enables visualization of the contribution of each stage to system dynamic range, identifying which stage or stages dominate performance. For example, the initial contributions for the example used in the case study that follows are shown in Table 7-1. The noise and distortion contributions indicate the stage that tends to have the most impact on noise factor or distortion. Redistribution of gain, noise, or distortion alters this relationship and leads to changes in system SFDR. Passive stages generally do not introduce distortion, but they do contribute noise. So, to simplify the following discussion, passive stages immediately preceding an active stage and the active stage itself will be considered as a single stage.

First of all, consider the traditional budget method approach in which the noise figure of a receiver cascade is established by maximizing the gain of the first stage, and thus the first stage establishes the SNR of the system since the noise contributions of subsequent stages are assumed to be negligible. With the system SNR fixed, a target system SFDR determines the overall IP3 performance described by the total system IIP3, IIP3^{T}. Also, in the budget method, the first stage has negligible impact on overall IIP3. Then a reasonable choice in the design process is to select stages following the first as contributing equally to the reduction in IIP3^{T}. Thus, in the budget method, the minimum acceptable IIP3 (in dBm) of the jth stage (i.e., $\text{IIP3}_{\text{dBm},j}$) in an m-stage cascade that is required to meet the target total IIP3 in dBm

Stage (i)	Gain (dB)	Gain G_i	NF_i (dB)	F_i	$IIP3_i$ (dBm)
1 filter	−3.98	0.4	3.98	2.5	–
2 LNA	1.76	1.5	3.01	2.0	-0.969
3 filter	−3.98	0.4	3.98	2.5	–
4 mixer	–	–	2.04	1.6	-3.18

Table 7-1: Stage assignments based on balanced contributions to the SFDR of the cascade. Initial assignment, $i = 1$.

($\overline{\text{IIP3}}^{\text{T}}_{\text{dBm}}$) is obtained from Equations (7.29) and (7.30) as

$$\text{IIP3}_{\text{dBm},j} = \text{IIP3}^{\text{A}}_{\text{dBm},j} + G^{\text{A}}_{\text{dB},(j-1)}, \tag{7.32}$$

where $$\text{IIP3}^{\text{A}}_{\text{dBm},j} = \overline{\text{IIP3}}^{\text{T}}_{\text{dBm}} + 10\log m. \tag{7.33}$$

(Note that the overline in $\overline{\text{IIP3}}^{\text{T}}_{\text{dBm}}$ identifies the target $\text{IIP3}^{\text{T}}_{\text{dBm}}$ and not the actual IIP3.) In addition, the gain or loss of preceding stages will modify this IIP3 as indicated in Equation (7.29). That is, with this assignment each stage makes an equal contribution to the overall intermodulation distortion. That is, each stage has the same IP3 referred to the system input. For example, in a cascade of three stages, each stage would need to have a minimum system IIP3 of 4.8 dB ($= 10\log 3$) in excess of the target system IIP3 (i.e., $\overline{\text{IIP3}}^{\text{T}}_{\text{dBm}}$).

If instead, using the contribution method, the cascade system was designed so that each stage contributed equally to the overall noise (C_j^N being the same for all stages), and each stage had the same gain (G_j being the same for all stages), then the required individual IIP3 values tend to be minimum. Maximizing system dynamic range becomes an exercise in maintaining the lowest noise power and highest IIP3 value throughout the cascade. The link between these parameters values is the distribution of gain and loss [7, 8].

For a transmitter, the SFDR is most commonly referred to the output and (repeating Equation (4.88))

$$\text{SFDR}_{\text{dB},o} = \tfrac{2}{3}\left(\text{OIP3}_{\text{dBm}} - N_{\text{dBm},o}\right), \tag{7.34}$$

and $\text{SFDR}_i = \text{SFDR}_o$. In the above, N_i and N_o are the total input and output noise powers and are assigned a noise floor value dependent on the cascade noise factor target and linear system gain.

7.4.4 Summary

The contribution method for designing systems of cascaded modules focuses on assigning the same dynamic range to each module. Of course linear modules such as a filter easily meet any system dynamic range requirement so the dynamic range assignments should focus on the nonlinear modules. Very often the dynamic range of an active module can be increased by either increasing the biasing of the module or switching to an alternative module but with high power requirements. Sometimes modules with different technology could provide increased dynamic range without increasing power consumption but such modules could be more expensive. The contribution method and the previously considered budget method only provide an initial starting point for system design. The system must still be optimized and often manually optimized as there are too many hard to quantify design goals. For example minimizing time-to-market and design cost are goals that cannot be parameterized.

7.5 Case Study: High Dynamic Range Down-Converter Design

In this section a down-converter in a receiver is designed. With the architecture and modules chosen in the RF cascade, the only quantities

that can be selected by the user are the gains of the amplifier stages. If adjusting the gains is not sufficient to meet system objectives, then it would be necessary to choose other modules or change the architecture.

7.5.1 Architecture

A down-converter with the architecture of Figure 7-8 is designed here considering the first four stages only and using the balanced contribution method for the initial stage assignment. The system operates at 1.5 GHz with a first IF of 250 MHz and an IIP3 design target of 0 dBm (i.e., $\overline{\mathrm{IIP3}}^{\mathrm{T}}_{\mathrm{dBm}} = 0\ \mathrm{dBm}$) with a noise figure target of 10 dB (i.e., $\overline{\mathrm{NF}}^{\mathrm{T}} = 10\ \mathrm{dB}$). The initial stage assignments for balanced noise and IMD contributions are shown in Table 7-1.

The stage assignments shown in Table 7-1 meet the target specifications exactly, with no single stage contributing any more than the required noise factor or nonlinearity (distortion) to meet the target. The LNA gain is low, permitting a significant reduction in required mixer IP3. In an RF system usually the mixer has the limiting distortion performance so anything that can be done to reduce the linearity required for a mixer is advantageous. The noise figure of a mixer can also be high. With this in mind, the process of system cascade trade-off and selection of actual element parameters continues in the following sections. That is, assignment of balanced contributions to each stage is a good initial starting point. This is followed up with further optimization and trade-offs.

7.5.2 Design

Design proceeds with a selection of available modules for stages and continuous but limited variation of stage parameters (through bias control, for example). The following stages were chosen for the stages (see Figure 7-8): an MMIC amplifier for Stage 2 (NEC part UPC2745) and a MMIC mixer for Stage 4 (MCL part SYM-2500). Bandpass dielectric resonator filters were chosen for Stages 1 and 3. In particular, a tuned lumped-element Chebychev bandpass filter with 0.1 dB ripple at 1.5 GHz was used to provide settable loss in Stage 3 and this proved to be important in establishing balanced stage contributions and improved dynamic range. The continuous control parameters include controlling the mixer IMD contribution by changing the LNA gain (through bias control), changing the mixer IIP3 (by varying the LO drive level), and retuning the Stage 3 filter. The noise and distortion contributions of stages with the initial stage assignment (from Table 7-1), initial measured performance, and optimized design are shown in Figures 7-9 and 7-10. Figure 7-11 depicts the dynamic range. Based on the balanced contribution method, the SFDR is 109.5 dB normalized to a 1 Hz bandwidth. The initial value provided by the devices selected is 108 dB. After appropriate trade-off and adjustable loss in Stage 3, the dynamic range achieved is 111 dB. Now the same system design could be achieved using the budget assignment initially, but with much greater design effort.

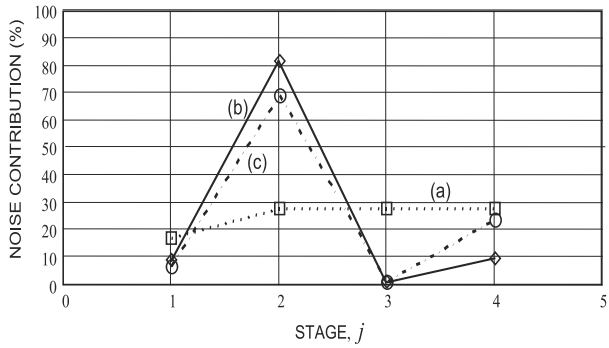


Figure 7-9: Noise figure contribution: (a) initial calculation of balanced noise figure contribution; (b) measured contributions of each stage using the initial assignments; and (c) final measured contributions after altering bias, filter loss, and mixer LO drive levels for improved dynamic range.

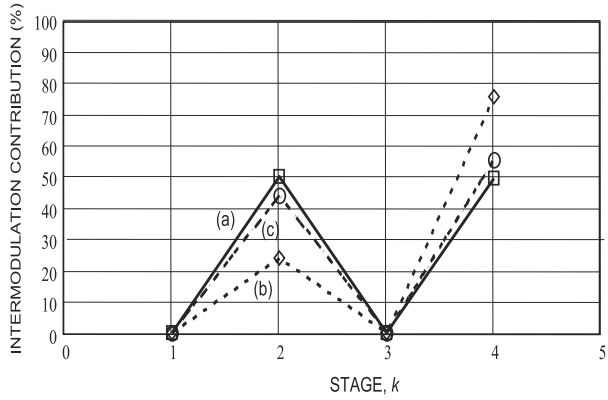


Figure 7-10: Input intercept (IP3) contribution: (a) initial calculation; (b) original measured contributions; and (c) final performance after alteration of bias, LO drive, and filter loss in Stage 3. Final settings in (c) led to the highest system dynamic range.

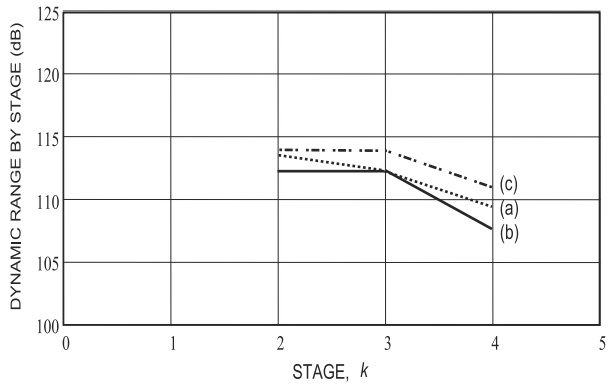
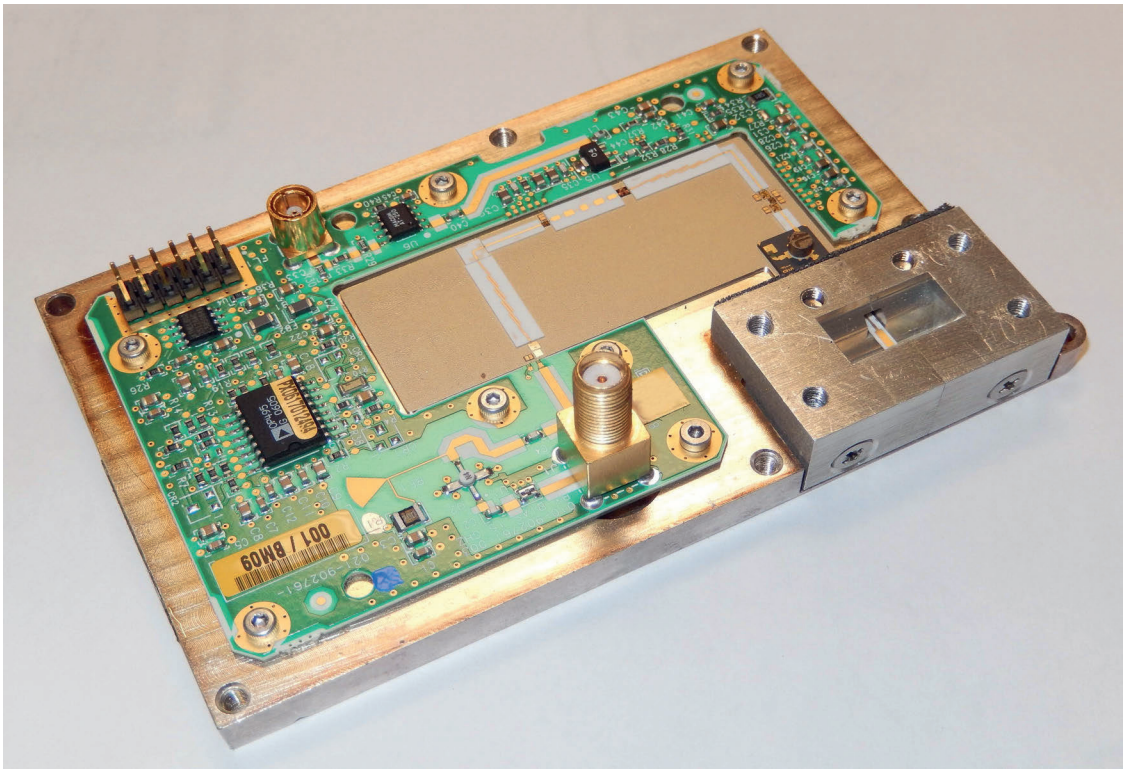


Figure 7-11: SFDR by stage: (a) initial assignment; (b) initial system performance after selection of modules but before optimization; and (c) final performance after setting bias, LO drive, and filter loss in Stage 3.

7.6 Case Study: Analysis of a 15 GHz Receiver

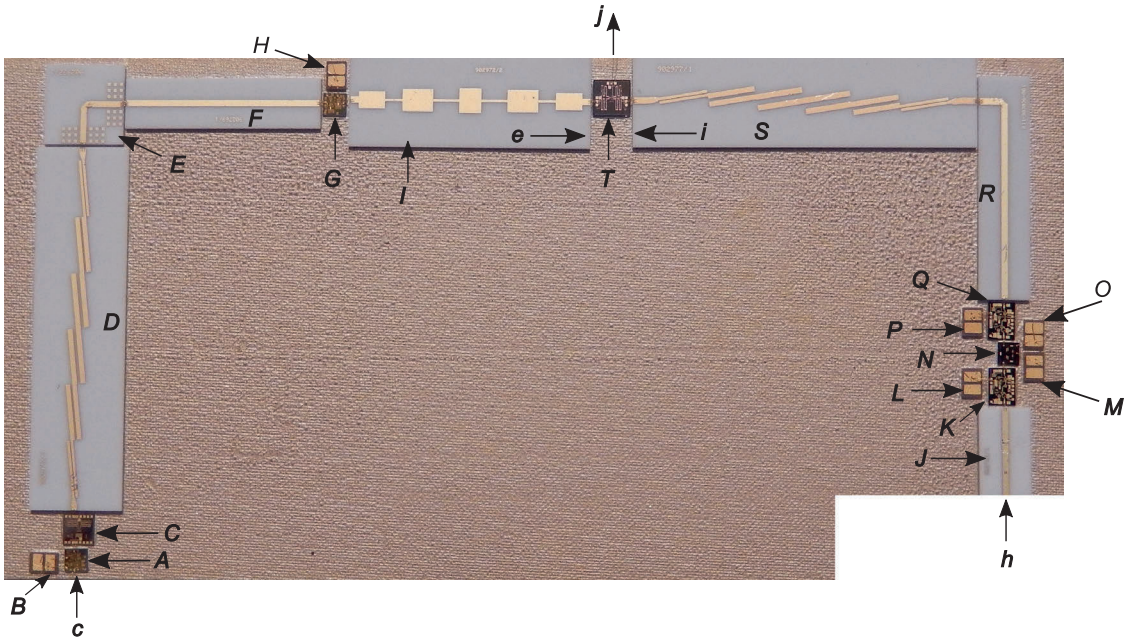
Most RF and microwave systems convert information at one frequency to information at another frequency that can either be more conveniently processed in the case of a receiver or more conveniently radiated in the case of a transmitter. Figure 7-12 is a 15 GHz receiver module which itself consists of interconnected sub modules. Detail of the frequency conversion section is given in Figure 7-13.

These sub modules could be called modules as well so that modules are constructed by interconnecting other modules many of which are available ”off-the-shelf” from high-level modules such as a receiver. The receiver module is part of a wireless service used in point-to-point microwave link such as in a cellular systems to provide communication between



Modules		Signal Path	
A	SMA connector, reference LO in	**a**	reference LO signal
B	Attenuator	**b**	Reference LO
C	Coupling capacitor	**c**	Reference LO
D	Reference LO Amplifier	**d**	LO with bandpass filter
E	Bias line	**e**	LO with lowpass filter
F	Radial stub	**f**	Radio frequency (RF) input signal (~15 GHz)
G	RF coupling capacitor (DC block)	**g**	RF in
H	Die attach mat	**h**	RF
I	Frequency multiplier	**i**	RF path with bandpass filter
J	Frequency multiplier	**j**	IF path with resistive antenuator
K	Waveguide-to-microstrip adaptor	**k**	IF path
L	Isolator	**l**	Intermediate frequency (IF) output
M	variable RF amplifier block		
N	Mixer		
O	IF amplifier		
P	Voltage variable attenuator		
Q	IF out, SMC connector		

Figure 7-12: A 14.4–15.35 GHz receiver module itself consisting of cascaded modules interconnected by microstrip transmission lines. Surrounding the microwave circuit are DC conditioning and control circuitry. Detail of the frequency conversion section mounted on the mat is shown in Figure 7-13.



A	Reference LO amplifier	*K*	RF amplifier
B	Power supply decoupling capacitor	*L*	Power supply decoupling capacitor
C	×2 frequency multiplier	*M*	Power supply decoupling capacitor
D	Edge-coupled PCL bandpass filter	*N*	Voltage variable attenuator
E	Microstrip bend	*O*	Power supply decoupling capacitor
F	Microstrip transmission line	*P*	Power supply decoupling capacitor
G	×2 frequency multiplier	*Q*	RF amplifier
H	Power supply decoupling capacitor	*R*	Microstrip transmission line and bend
I	stepped impedance lowpass filter	*S*	Edge-coupled PCL bandpass filter
J	Microstrip transmission line	*T*	Subharmonic mixer

Figure 7-13: Frequency conversion section of the receiver module shown in Figure 7-12. The reference LO is applied to the frequency conversion section at *c*, the RF is applied at *h* following the isolator. The IF is output at *j*.

basestations. The subsystem modules such as the amplifiers, frequency multipliers, mixers, isolator, and waveguide adapter are available as off-the-shelf components from companies that specialize in developing such modules and selling them to a large user base. The block diagram of the receiver module is shown in Figure 7-14.

The RF module engineer must interconnect these modules and mange interference and dynamic range. The RF module designer either designs the various filters or specifies them for design and fabrication by specialty microwave filter companies. The conductive die attach mat **H** is material silk-screened on to the brass back-plate and enables the alumina interconnects, and the amplifier and mixer die, to be epoxy-attached. It is not wise to directly attach die and ceramic substrates directly to the brass housing as the difference in thermal coefficients of expansion and the rigidity of the two systems means that the attachment could fail or the ceramic and die crack. The mat is said to be 'compliant' providing stress relief when the solid

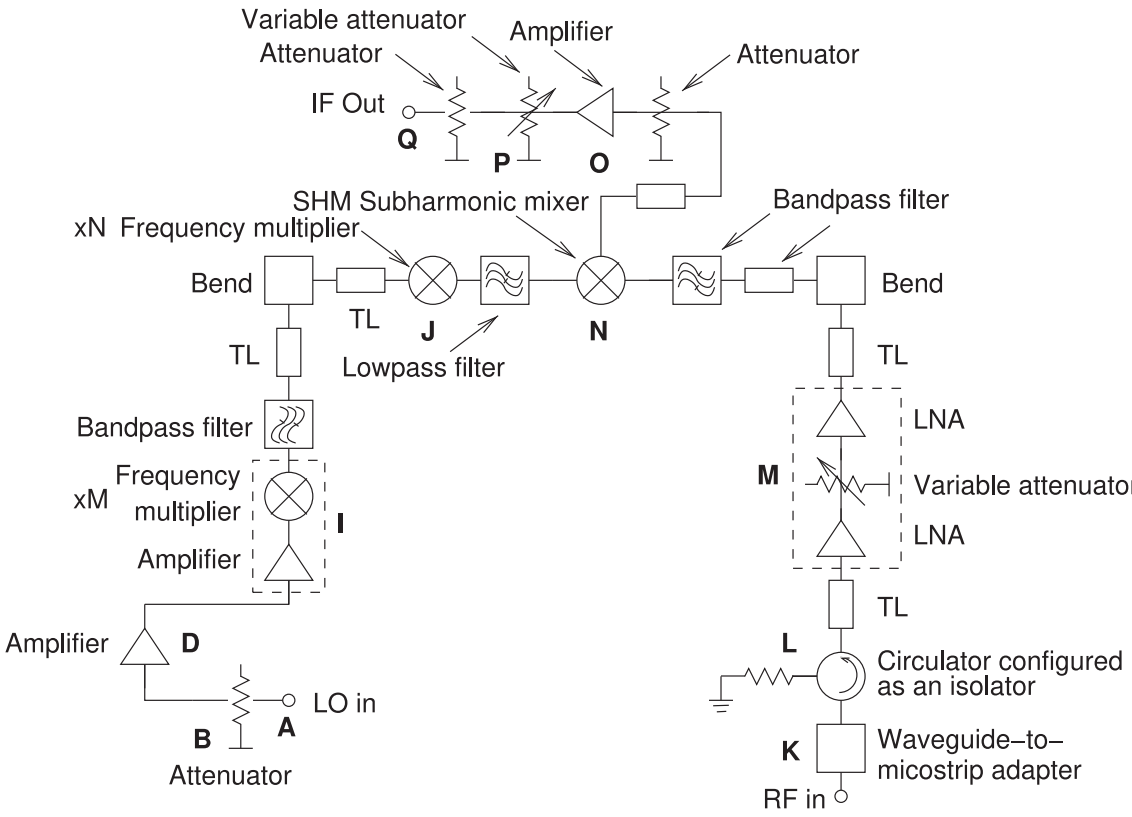


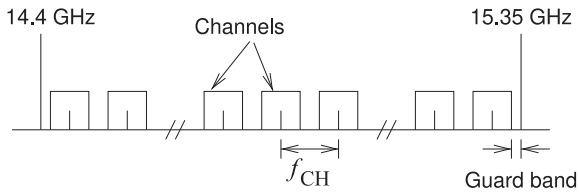
Figure 7-14: Block diagram of the receiver module shown in Figure 7-12.

objects expand at different rates. The ceramic substrates, usually alumina, have repeatable and stable dimensions for realization of filters which have a response that strongly depends on dimensions.

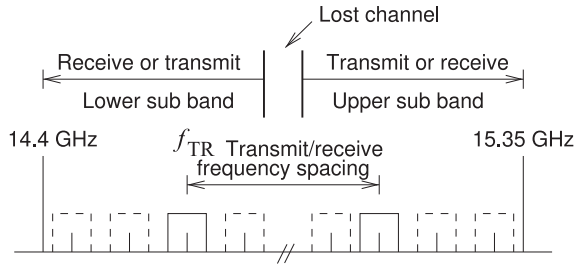
7.7 Case Study: Frequency Planning of a Transceiver

This case study develops the architecture and frequency plan of a transceiver for the licensed fixed wireless service in the 15 GHz band that provides channels supporting up to 140 Mbit/s. This is the service that provides two-way point-to-point communications most commonly to connect pairs of cellular basestations to each other. A central basestation in a cluster of basestations then connects traffic into a backbone fiber optic-based network. Frequency planning begins with the specifications provided by national regulatory authorities which generally adopt recommendations from the International Telecommunications Union. (See [9] for the recommendations concerning the 15 GHz fixed wireless service.)

The 15 GHz fixed wireless band extends from 14.4 GHz to 15.35 GHz supporting a number of channels with various channel spacing. The spacing used will be based on local regulations and licensing. A universal transceiver must support channel spacings of 2.5, 3.5, 7, 14, 28, and 56 MHz [9]. Figure 7-15(a) shows the channel assignment where f_{CH} is the channel spacing. No



(a) Fixed service channels



(b) Transmit/receive channel assignment

Figure 7-15: Channel assignment in the 15 GHz fixed wireless service band.

.

guard bands between the channels are specified but typically a few 10's of kilohertz are sufficient. Thus channel spacing is only slightly larger than the modulation bandwidth and so there is a narrow guard band. Such a relatively tight guard band is possible as significant digital processing power can be used to resolve adjacent channels. At the low and high ends of the 15 GHz band there is a guard band that is about half of a channel spacing so that other wireless services are not affected.

A transceiver in the fixed wireless service supports one or more transmit and receive channel pairs. As specification in the ITU regulations the separation of transmit and receive channel pairings can be 315, 322, 420, 490, 640, 644, 728, 840 MHz. That is, if a receive (or transmit) channel is centered at 14.500 GHz, then the paired transmit(or receive) channel is centered at 14.815, 14.822, 14.920, 14.950, 15.140, 15.144, 15.228, or 15.340 MHz. Since a two-way link is established, the center frequency of the received channel at one end of the link is the center frequency of the transmitted channel at the other end of the link.

7.7.1 Transceiver Architecture

A transceiver must simultaneously transmit and receive channels and usually through the same antenna. Even if separate transmit and receive antennas are used, there will be significant coupling of the transmitted and received signal if they are located near each other. In practice transceivers use a diplex filter, a diplexer, to separate the 15 GHz band into two sub bands of equal bandwidth. In principal a diplex filter can be two filters combined to provide a lowpass function for the upper sub band and a highpass filter for the lower sub band. For fixed wireless service operation the diplex filter is designed as two adjacent bandpass filters, one passing the lower sub band and the other passing the upper sub band. Although the filter skirts can be designed to be very steep, they necessarily have finite frequency extent so

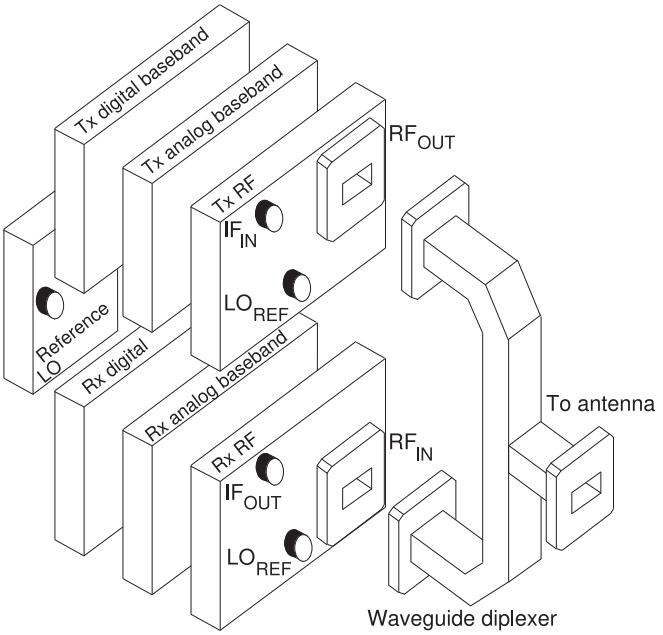


Figure 7-16: Transceiver architecture.

that the middle of the channel is no longer available to this transceiver.

The high-level architecture of a transceiver for the 15 GHz fixed wireless service is shown in Figure 7-16. This architecture provides high isolation of the transmit and receive paths and supports a modular design approach. The transmit section comprises the three modules shown on the top of the figure. The first module is the transmit (Tx) digital baseband module that prepares the digitally modulated signal for one (or more) transmit channels. Either Reed-Solomon or trellis coded modulation (TCM) forward error correction schemes are used. Five modulation formats are typically supported: QPSK, QAM16, QAM32, QAM64 or QAM128 with the modulation scheme adjusted to maintain a maximum bit error rate of 10^{-6}. The signal in digital form is based in the transmit digital baseband module. The transmit analog baseband module modulates the signal on a carrier at an intermediate frequency of around 1 GHz. This is then passed to the transmit RF module which generates the RF signal in the 15 GHz band. Typically the transmitted power is 26 dBm using QPSK modulation, 23 dBm using QAM16, 22 dBm using QAM32, 20 dBM using QAM64, and 19 dBm using QAM128 modulation. The signal then passes into one port of the diplexer, here a waveguide diplexer, before passing to the antenna. The receiver sensitivity is typically -93 dBm for QPSK, -86.5 dBm for QAM16, -75 dBm for QAM32, -73.5 dBm for QAM64 modulation, and -68.5 dBm for QAM128 modulation, again for a BER of better than 10^{-6}.

The waveguide diplexer is shown again in Figure 7-17. High performance diplexers can be designed in waveguide as there is very little loss and the Qs of the filter resonators are high [10, 11]. Here the waveguide diplexer has a bandpass response that passes the lower sub band, and a second bandpass response that passes the upper sub band. The sub band responses are shown in Figure 7-15(b). The two bandpass responses are designed together so that the diplexer performance is better than if two independently designed bandpass filters were used. Either sub band output of the diplexer could be

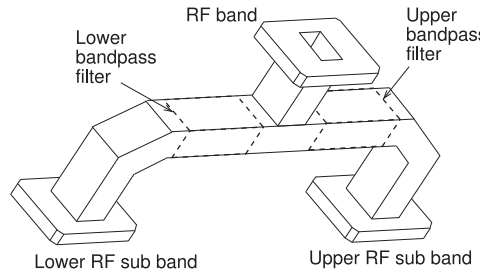


Figure 7-17: Waveguide diplexer with two bandpass filters inside the waveguide sections. One allows the lower RF sub band to pass and the other lets the upper sub band pass. The waveguide filter transition can be very sharp so that the edges of the upper and lower sub bands can be very close eliminating only one channel.

connected to the transmitter (and the other to the receiver) depending on whether the channel transmitted is in the lower or upper sub band.

The bottom three modules shown in Figure 7-16 comprise the receiver. The RF signal enters the Receiver (Rx) RF module at $\mathrm{RF_{IN}}$ where it is down converted to an intermediate frequency signal around 1 GHz. The intermediate signal passes to the Receive Analog Baseband module where the desired channel (or perhaps more than one) is extracted and presented to the receive digital baseband module at a frequency less than 100 MHz.

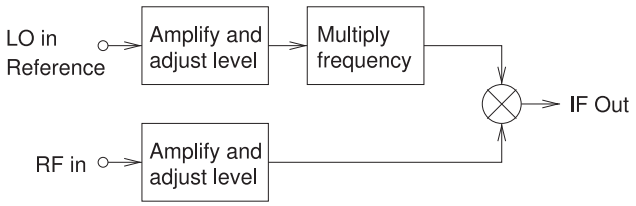
A common LO source is generated in the Reference LO module with one output connected by cables to the Receive RF module and another version of the LO, shifted by the appropriate receive/transmit frequency spacing, routed to the Transmit RF module.

The transmit and receiver RF and analog baseband modules must be mounted on the communication tower but the digital baseband and reference LO modules can be located in an air conditioned room. Then coaxial cables with signals up to 2 GHz route the RF signals and the reference LO signals between the ground unit and the tower-mounted unit.

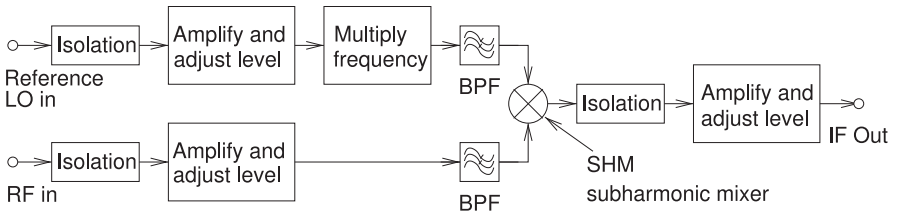
7.7.2 Frequency Planning

Basic frequency planning began with definition of the basic transceiver architecture with the choice of an intermediate frequency after first conversion (or before the final conversion in the case of the transmitter) that is around 1 GHz. This choice was made as the transceiver must be able to transmit across the entire 950 MHz bandwidth of the 15 GHz band. Once the orientation of the diplexer has been fixed before delivery to the customer, or in field installation, the transmit band is from one sub band to the other sub band, a narrower but still significant 475 MHz bandwidth. However the unit must still be designed to support operation over the whole band. Thus, for the receiver, the design decision has been made to down convert an entire sub band to a frequency range centered around 1 GHz. Such a circuit must have essentially a lowpass response but with signals at DC and up to a hundred megahertz or so above blocked by inductors and capacitors providing DC connections and/or RF isolation. The fractional bandwidth is too much for the design of a bandpass filter. The frequency conversion description is similar for the transmitter.

Frequency planning has a major impact on the design cost of the transceiver and in the unit cost, especially if extensive tuning is required during manufacture. Frequency planning requires considerable experience with knowledge of design cost drivers, manufacturability, which modules and sub modules are available from vendors, and what one's own organization can do to provide a competitive advantage. To be competitive



(a) Basic signal flow in the RF receiver



(b) Extended signal flow in the RF receiver

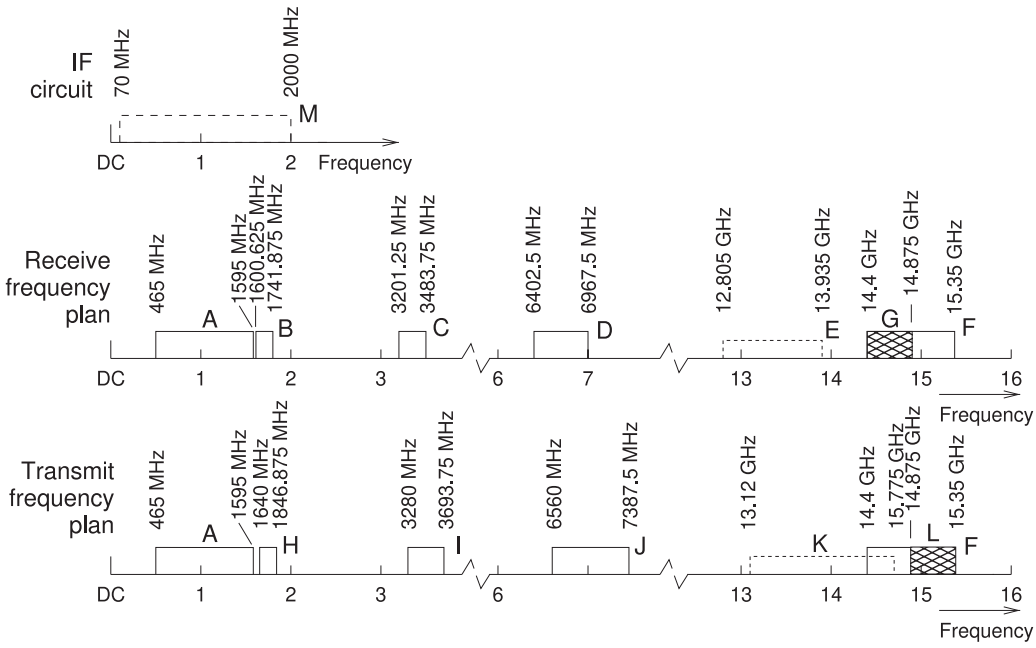
Figure 7-18: Signal paths in the receive RF module in Figure 7-16.

the transceiver must be tunable across the entire 15 GHz band.

Cost is substantially reduced by sharing an LO between the transmitter and receiver, with either the transmit or receiver versions of the LO shifted by the appropriate transmit/receive spacing. The LO will be locked to a low frequency reference using a phase-locked loop and then it must be routed by coaxial cable to the receive and transmit RF modules. This means that LO should ideally be below 2 GHz. This LO will now be called the reference LO as it is not the ultimate frequency that is required by the mixer in the frequency conversion section of the receiver RF module. Ultimately the effective LO at the mixer must be tunable across the entire 15 GHz band and so needs to tune at least by 950 MHz. Tuning can be accomplished using phase-locked loops but realistically the tuning range that can be conveniently achieved is 200 MHz or 15%, whichever is less up to 2 GHz.

The basic signal flow in the receiver RF module is shown in Figure 7-18(a). The basic concept is that the received RF signal is mixed with an LO to produce a low frequency version as the intermediate frequency signal, the IF. In the signal flow, a typically low-level RF signal is presented to the receiver at the port indicated by $\mathrm{RF_{IN}}$. The RF signal is amplified by a low noise amplifier so that circuit noise in the following RF signal path has no impact. Since the level of the received signal is not known, the amplitude of the signal must be leveled. The reference LO is at a frequency below that required to pump the mixer so that it can be more easily generated and routed in the transceiver. Generally it would be preferable to keep the signals to be routed below 2 GHz as then the quality of the cabling and connections is less critical. As a result the frequency of the reference LO signal must be increased before it is presented to the mixer.

Knowledge of which sub modules are readily available is also critical in frequency planning. The use of frequency multipliers, essentially nonlinear harmonic generators followed by filters, can increase the frequency by a factor of 8 or so quite reliably. With frequency multiplication there is an increase in the noise levels of the oscillator that impacts performance and



Receive frequency plan		Transmit frequency plan	
A	IF band	A	IF band
B	Rx reference LO	H	Reference LO
C	2nd harmonic of Rx reference LO	I	2nd harmonic of Tx reference LO
D	4th harmonic of Rx reference LO	J	4th harmonic of Tx reference LO
E	8th harmonic of Rx reference LO	K	8th harmonic of Rx reference LO
F	RF band	F	RF band
G	Lower RF sub band	L	Upper RF sub band

Figure 7-19: Transceiver frequency plan. The receive and transmit frequency plans are exchanged if the RF received is in the upper RF band (and then the RF transmitted is in the lower RF sub band. Also shown is the IF circuit bandwidth extending from 70 MHz to 2 GHz.

this means that the noise performance of the reference LO needs to be that much better. Furthermore there is a need to adjust signal levels to avoid distortion. One of the essential characteristics of most module design is that modules can be freely interconnected without concern that one module will affect the operation of another module. To achieve this isolation is required. Another design choice is to use a subharmonic mixer in which the the pump frequency is about half of the RF frequency. The subharmonic mixer also has the property that very little of the second harmonic of the LO signal passes into the RF circuitry. It is very important that the range of frequencies in the LO path, the RF path, and the IF path do not overlap. With this knowledge the signal flow path in the receiver RF module can be extended as shown in Figure 7-18(b).

Using the above considerations and also considering the transmit requirements leads to the detailed frequency plan shown in Figure 7-19. This plan is arrived at iteratively considering design and unit cost, as well as the required time to market. While Figure 7-19 shows the receiver using the lower sub band and the transmitter using the upper side bands, this sub band allocation could be reversed and this is achieved by simply switching

the diplexer sub band ports, i.e. by rotating the waveguide diplexer.

7.7.3 Summary

Once the frequency plan has been established the design of the individual modules can proceed independently. The final receiver RF module was shown in Figures 7-12 and 7-13 with the block diagram of the module shown in Figure 7-14.

Frequency planning may seem rather simple but it is more than drawing a spectral diagram such as that in Figure 7-19. The architecture of the transceiver is defined along with the frequency plan. Thus frequency planning is done by experienced engineers with a broad appreciation for RF system performance, knowledge of available off-the-shelf modules, appreciation for cost, and appreciation for what it takes to develop a competitive product. Frequency planning has a tremendous impact on system cost. The final system design has a value far exceeding the cost of the component sub modules.

7.8 Summary

A microwave system or subsystem is generally constructed as a cascade of two-port modules. Typically such modules are interconnected on a circuit board using microstrip structures with the system designer providing matching networks, transmission line networks, system architecture, and frequency plan. One or two decades ago crafting such a system using vendor-supplied modules would have required a significant performance compromise. As a result microwave companies necessarily developed large portions of a design in-house that resulted in long design cycles. With programmability and user-defined adjustments, perhaps by setting bias levels, the gain, bandwidth, and distortion performance of off-the-shelf modules can be adjusted. Active modules necessarily introduce distortion and noise, and so managing dynamic range while managing DC power consumption is a considerable concern.

Design and analysis of RF and microwave systems is complicated by the complex signals, i.e., modulated signals and not discrete sinewaves, used in RF and microwave systems and by the excessively long simulation times required to analyze microwave circuits with these signals. Design of linear RF and microwave subsystems such as an amplifier or matching network can proceed very well by simulating the structure one frequency at a time and then stepping the signal frequency over the range of interest.

The analysis of nonlinear microwave subsystems such as amplifiers is more complicated but often it is sufficient to consider two-tones, signals with two sinusoidal components and then there are efficient simulation techniques such as harmonic balance analysis that enables subsystem performance to be efficiently but approximately evaluated. However, neither of these strategies is sufficient when trying to determine the performance of RF and microwave systems. In this case complex signals such as digitally modulated signals must be considered, and there are usually signals at widely different frequencies to be incorporated in any analysis. There is however one special property of microwave signals that is exploited, and this is that nearly every modulated microwave signal has a relatively small

bandwidth compared to the center frequency of the signal. Thus system simulation strategies have been developed that require only the smaller bandwidth signal to be fully considered in analysis and bookkeeping used to track the center frequency of the signal.

In this chapter metrics and design techniques were introduced that enable competitive designs to be realized quickly. Often a company may decide that it is to their competitive advantage to design one or more of their own modules rather than using those from module vendors. Even then, the large number of modules available enables design concepts to be tested for feasibility early in the design cycle. Module vendors are challenged with designing modules with close to ultimate performance, but with enough adaptability that the module is suited to a wide variety of system applications. The greater customer base for modules justifies the higher module design costs leading to acceptable unit costs and high performance.

7.9 References

[1] H. Gutierrez, K. Gard, and M. Steer, "Nonlinear gain compression in microwave amplifiers using generalized power-series analysis and transformation of input statistics," *IEEE Trans. on Microwave Theory and Techniques*, vol. 48, no. 10, pp. 1774–1777, 2000.

[2] K. Gharaibeh, K. Gard, H. Gutierrez, and M. Steer, "The importance of nonlinear order in modeling intermodulation distortion and spectral regrowth," in *2002 IEEE Radio and Wireless Conf., 2002 (RAWCON 2002)*, 2002, pp. 161–164.

[3] S. Maas, "Third-order intermodulation distortion in cascaded stages," *IEEE Microwave and Guided Wave Letters*, vol. 5, no. 6, pp. 189–191, Jun. 1995.

[4] ——, *Nonlinear Microwave and RF Circuits*, 2nd ed. Artech House, 2003.

[5] R. Sagers, "Intercept point and undesired responses," *IEEE Trans. on Vehicular Technology*, vol. 32, no. 1, pp. 121–133, Feb. 1983.

[6] A. Victor and M. Steer, "Transceiver cascade system analysis and design via a contribution method," *Int. J. on RF and Microwave Computer Aided Engineering*, vol. 16, no. 4, pp. 338–345, Jul. 2006.

[7] R. Pettai, *Noise in Receiving Systems*. John Wiley & Sons, 1984.

[8] W. Sabin and E. O. Schoenike, *Single-Sideband Systems and Circuits*. McGraw-Hill, 1987.

[9] "Itu recommendation f.636 f.636 : Radio-frequency channel arrangements for fixed wireless systems operating in the 14.4-15.35 ghz band," Mar. 2012.

[10] A. Kirilenko, S. Senkevich, V. Tkachenko, and B. Tysik, "Waveguide diplexer and multiplexer design," *IEEE Trans. on Microwave Theory and Techniques*, vol. 42, no. 7, pp. 1393–1396, Jul. 1994.

[11] S. Shin and S. Kanamaluru, "Diplexer design using EM and circuit simulation techniques," *IEEE Microwave Magazine*, vol. 8, no. 2, pp. 77–82, Feb. 2007.

7.10 Exercises

1. The first stage of a two-stage amplifier has a linear gain of 16 dB and output 1 dB gain compression, $P_{1o,1\text{dB}} = -20$ dBm. For the second stage the linear gain is 30 dB and $P_{2o,1\text{dB}} = 0$ dBm.
   (a) Determine the input-referred gain compression, $P_{2i,1\text{dB}}$ of stage 2.
   (b) Compare $P_{2i,1\text{dB}}$ and $P_{1o,1\text{dB}}$. Which stage dominates gain compression?
   (c) What is the amplifier's output gain compression level, $P^T_{o,1\text{dB}}$ considering only compression from the dominant stage.
   (d) Calculate $P^T_{o,1\text{dB}}$ using the method described in Section 7.2.1.
   (e) Compare $P^T_{o,1\text{dB}}$ calculated in (c) and (d) and briefly discuss any discrepancy.
2. An amplifier has two cascaded stages with linear gains of $G_1 = 20$ dB and $G_2 = 30$ dB, and output-referred third-order intercepts of $\text{OIP3}_1 = 0$ dBm and $\text{OIP3}_2 = 20$ dBm, respectively. What is IIP3 of the amplifier? Use the organized cascade intercept method.
3. A single-stage amplifier has a linear gain of 16 dB and an output 1 dB gain compression point of 10 dBm. A communication signal with

a PMEPR of 6 dB is used. What is the maximum average power of the input signal before the output suffers significant compression? This is defined at the point at which the peak signal is compressed by 1 dB.

4. The first stage of a two-stage amplifier has a linear gain $G_1 = 30$ dB and an output 1 dB gain compression point $P_{1o,1\text{dB}} = -10$ dBm. The second stage has a linear gain $G_2 = 20$ dB and an output 1 dB gain compression point $P_{2o,1\text{dB}} = 10$ dBm. What is the output-referred 1 dB gain compression point of the cascade amplifier? [Parallels Example 7.1]
5. An amplifier consists of two cascaded stages. The first stage has a linear gain $G_1 = 30$ dB and an output 1 dB gain compression point $P_{1o,1\text{dB}} = 0.1$ dBm. The second stage has a linear gain $G_2 = 20$ dB and an output 1 dB gain compression point $P_{2o,1\text{dB}} = 1$ dBm. What is the input-referred 1 dB gain compression point of the cascade amplifier? [Parallels Example 7.1]
6. The stages of a two-stage amplifier have linear gains of G_1 and G_2, and output 1 dB gain compression powers of $P_{1o,1\text{dB}}$ and $P_{2o,1\text{dB}}$, respectively. Develop a symbolic expression for the input-referred 1 dB gain compression point of the cascade amplifier.
7. An amplifier has two stages with linear gains of $G_1 = 20$ dB and $G_2 = 30$ dB, and output 1 dB gain compression powers of $P_{1o,1\text{dB}} = 0.1$ dBmand $P_{2o,1\text{dB}} = 1$ dBm, respectively. What is the input-referred 1 dB gain compression power of the amplifier?
8. The first stage of a two-stage amplifier has a linear power gain of 26 dB and an output 1 dB gain compression power of 10 dBm. The corresponding parameters of the second stage are 10 dB and 13 dBm.
   (a) What is the linear power gain of the two-stage amplifier?
   (b) What is the output 1-dB gain compression power of the amplifier for a sinusoidal RF input signal?
   (c) What is the maximum average output RF power of the 64-QAM-modulated signal (with a PMEPR of 7.8 dB) for an undistorted output (as defined by 1-dB gain compression)?
9. The stages of a three-stage amplifier have linear gains of 10 dB, 20 dB, and 20 dB respectively, and 1 dB output gain compression levels of -60 dBm, -40 dBm, and -20 dBm respectively. What is the output power when the gain of the amplifier is compressed by 1 dB?
10. The first stage of a two-stage amplifier has a linear power gain of 26 dB, an output 1 dB gain compression power of 10 dBm, and an output-referred third-order intercept point OIP3 = 26 dBm. The second stage has a linear power gain of 10 dB, an output 1 dB gain compression point of 13 dBm, and an output-referred third-order intercept point OIP3 = 33 dBm.
    (a) What is the linear power gain of the two-stage amplifier?
    (b) What is the output 1-dB gain compression power of the two-stage amplifier for a sinusoidal RF input signal?
    (c) What is the OIP3 of the two-stage amplifier?
    (d) What is the input-referred third-order intercept point, IIP3?
11. The final RF output of a cell phone has a driver amplifier followed by a power amplifier. The driver amplifier has a linear gain of 30 dB and an output-referred third-order intercept point, OIP3, of 50 dBm. The power amplifier has a linear gain of 12 dB and an output-referred third-order intercept point, OIP3, of 55 dBm. What is the OIP3 of the driver-power amplifier cascade?
12. A two-stage amplifier has a linear power gain of 20 dB, an output 1 dB gain compression point of 30 dBm, and an output-referred third-order intercept point OIP3 = 53 dBm.
    (a) What is the power of the maximum input signal when the gain of the amplifier is compressed by 1 dB?
    (b) What is the input-referred third-order intercept point, IIP3?
13. The first stage of a two-stage amplifier has a linear power gain of 23 dB, an output 1 dB gain compression power of 1 dBm, and an output-referred third-order intercept point OIP3 = 20 dBm. The second stage has a linear power gain of 10 dB, an output 1-dB gain compression point of 10 dBm, and an output-referred third-order intercept point, OIP3 = 30 dBm. The signal applied to the amplifier uses QPSK modulation with a PMEPR of 3 dB.
    (a) What is the linear power gain in decibels of the two-stage amplifier?
    (b) What is the output 1-dB gain compression power, in dBm, of the two-stage amplifier?
    (c) What is the OIP3, in dBm, of the two-stage amplifier?
    (d) What is the input-referred third-order intercept point, IIP3m?
    (e) What is the single-tone output power at 1 dB gain compression?
    (f) What is the maximum output RF power of

the QPSK-modulated signal for an undistorted output?

14. An amplifier has two cascaded stages. The stages have linear gains of G_1 and G_2, and output-referred third-order intercepts of $\mathrm{OIP3}_1$ and OIP3 dBm, respectively. What is IIP3 of the amplifier?

15. An amplifier has two cascaded stages with linear gains of $G_1 = 20$ dB and $G_2 = 30$ dB, and output-referred third-order intercepts of $\mathrm{OIP3}_1 = 0$ dBm and $\mathrm{OIP3}_2 = 20$ dBm, respectively. What is IIP3 of the amplifier? Use the unorganized cascade intercept method.

16. The first stage of a room-temperature two-stage amplifier with a 100 MHz bandwidth has a linear power gain of 26 dB, an output 1 dB gain compression power of 10 dBm, and an output-referred third-order intercept point OIP3 = 26 dBm. The second stage has a linear power gain of 6 dB, an output 1 dB gain compression point of 13 dBm, and an OIP3 of 33 dBm. The noise figure of the first stage is 3 dB and the noise figure of the second stage is 6 dB. The minimum acceptable SNR, $\mathrm{SNR}_{\mathrm{min}}$, at the output of the amplifier is 16 dB.
    (a) What is the linear power gain of the two-stage amplifier?
    (b) What is the output 1-dB gain compression power of the two-stage amplifier for a sinusoidal RF input signal?
    (c) What is the OIP3 of the two-stage amplifier?
    (d) What is the noise figure of the two-stage amplifier?
    (e) What is the noise, in dBm, applied to the input of the two-stage amplifier in a 100 MHz bandwidth is the source has a Thevenin resistor at room temperature?
    (f) What is the power of the noise, in dBm, in a 100 MHz bandwidth at the output of the two-stage amplifier?
    (g) What is the output-referred spurious free dynamic range of the two-stage amplifier in decibels?
    (h) What is the output-referred dynamic range of the two-stage amplifier in decibels?

7.10.1 Exercises By Section

†challenging, ‡very challenging

§7.2 $1^{\dagger}, 2, 3, 4, 5^{\dagger}, 6^{\dagger}, 7^{\dagger}, 8, 9, 10, 11^{\dagger}, 12^{\dagger}, 13^{\dagger}, 14, 15$

§7.7 $16^{\dagger}$

7.10.2 Answers to Selected Exercises

3 -11 dBm

10(e) 5.2 dBm

Index

A/D converter, 186
accumulator, 213
active
  filter, 64, 66
ADC, 186
admittance
  inverter, 33, 37
all pole, 18
allpass, 12
AM-AM, 144
AM-PM, 144
amplifier
  back-off, 145
  operational, 67
  transconductance, 67
amplitude
  demodulator, 212
  distortion, 144
analog
  -to-digital converter, 186
  multiplier, 198, 206

B, 131
back-off
  gain, 144
backward-wave oscillator, 216
balanced
  mixer, 197
bandpass filter, 12
  transient response, 69
bandstop filter, 12
BER, 153
Bessel, 17
biquad filter, 67
biquadratic filter, 67
BJT
  schematic symbol, 8
Boltzmann constant, 131
breakdown, 156
brick-wall filter, 17
budget method, 231, 232
Butterworth, 17
  filter, 18
BWO, 216

capture range, PLL, 210
carrier
  oscillator, 203
cascade
  budget method, 231
  contribution method, 232
  IIP3, 228, 234
  intercept method, 228
  noise, 133, 233
  OIP3, 228, 234
  organized intercept method, 229
  unorganized intercept method, 230
case study
  bandstop filter, 62
  cascade design, 236
  combline filter, 93
  down converter, 236
Cauer, 17, 29, 31, 44–46
characteristic
  function, 15
  polynomial, 20
Chebyshev, 17, 23
  filter, 22
chirp, 69
circulator, 177, 179
coherent
  amplitude demodulator, 212
combline
  filter, 93
    case study, 93
commensurate frequency, 40, 97, 99
commutating mixer, 195, 197
compression
  gain, 145
contribution method, 232
conversion, 186
  down, 186
  loss, 189
  up-conversion, 186
corner frequency, 12
corona effect, 157
correlation
  noise, 128

DAC, 213
DDS, 213
demodulator, amplitude, 212
detectable signal
  minimum, 152
Dicke switch, 139
dielectric
  resonator, 62
digital-to-analog converter, 213
diode, 167
  Gunn, 215
  IMPATT, 215
  junction, 167
  light-emitting, 169
  mixer, 187
  PIN, 169
  Schottky, 167
  transferred electron device, 215
  transit time device, 215
  tunnel, 214
  TUNNETT, 215
  varactor, 168
  Zenner, 169
direct
  digital synthesizer, 213
discernible signal, minimum, 152
distortion, 144, 225
  AM-AM, 144
  AM-PM, 144
  amplitude, 144
  passive intermodulation, 154
  phase, 144
  PIM, 154
divider
  frequency, 208
double
  balanced
    mixer, 197
  sideband, 190
doubly terminated
    network, 13, 14
down-conversion, 186
DPDT switch, 175
DPST switch, 175
DR, 151–153
duplexer
  radar, 180
DUT, 154
dyadic, 178
dynamic
  range, 151
    cascaded system, 235
    spurious free, 151

effective noise temperature, 128
electron spin, 177
equiripple, 23
excess noise, 131

F, 131
Fathelbab, 11
ferrite, 177
FET, 175
  switch, 175
field
  effect transistor, 175
figure of merit
  oscillator, 204, 205
filter, 11, 79
  design using image parameter method, 13
  design using insertion loss method, 13, 14
  design using reflection coefficient method, 13, 14
  active, 64, 66, 67
    distributed, 68
  allpass, 12
  bandpass, 12, 61
  bandstop, 12, 62
  Bessel, 33
  biquad, 67
  biquadratic, 67
  Butterworth, 30, 32
    prototypes, 29
  Cauer, 32
  Chebyshev, 22, 31, 32
    prototypes, 29
    Type I, 33
    Type II, 33
  combline, 93
    Chebyshev, 93
  corner frequency, 40
  dielectric resonator, 62
  elliptical, 32
  end-coupled, 61
  equiripple, 23
  frequency
    scaling, 40, 41
    transformation, 41
  group delay, 60
  highpass, 12
  impedance scaling, 40
  impedance transformation, 40
  inverter
    admittance, 37, 38
    lumped-element, 101

scaling, 91
ladder, 46
Legendre, 33
lowpass, 12
to bandpass, 42, 43
to bandstop, 43, 45
to highpass, 43
optimum "L", 33
parallel coupled line, 61, 79
PCL, 61, 79
RFIC, 66
Richards' transformation
lowpass, 52
Richards's
transformation, 50, 53
highpass, 53
synthesis, 18
tank, 68
transformations, 40
transient response, 69
type transformation, 40
YIG, 179
YIG tuned, 181
final-value theorem, 211
first
order loop, 210
flicker noise, 202
FOM
oscillator, 204, 205
FPD6836P70
noise parameters, 141
frequency
commensurate, 97
conversion, *see* conversion
demodulator, PLL, 211
divider, 208
locked-oscillator, 209
down-conversion, *see* down-conversion
heterodyne, *see* heterodyne
multiplier, 207
parametric, 208
reactive, 208
resistive, 207
parametric
multiplier, 208
reactive multiplier, 208
resistive multiplier, 207
scaling (filters), 41
synthesizer, 211
transformation (filters), 41
up-conversion, *see* up-conversion
Friis's formula, 134, 233

gain
compression, 145
back-off, 144
cascaded system, 225
Gilbert
cell, 198
mixer, 198
group
delay, 60
filter, 60
Gunn
diode, 215
effect device, 215
gyromagnetic
effect, 177, 181
resonance, 178

HBT
schematic symbol, 8
heterodyne, 186
heterodyning, 186
highpass, 12
Hurwitz polynomial, 17

IF, 186
IIP3, 152
cascaded system, 228, 234
IL, 14
IM2, 151
IM3, 146, 148, 153
image
parameter method, 13
reject mixer, 197
rejection, 190
ratio, 190
IMP, 146
IMPATT diode, 215
impedance
inverter, 33, 34
scaling, 40
transformation, 40
implementation margin, 152
inhomogeneous
medium
parallel coupled-line filter, 114
PCL filter, 114
insertion loss, 14
intercept
IP3, 152
method, 152, 228–230
intermod, 146
intermodulation
distortion
cascaded system, 228, 234
product, 146
inverter, 33
admittance, 33, 37, 38
impedance, 33, 34, 36
lumped-element, 101
scaling, 91
stub realization, 39
IP2, 225
IP3, 225
intercept, 152
IRR, 190
isolator, 177, 180

j, 21
junction diode, 167

k, 131
klystron, 216
Kuroda's identities, 54

L_C, 189
lanthanoid elements, 177
LC resonator
stub equivalence, 95
LED, 169
Leeson, 204
Levy, 27
light-emitting diode, 169
linear chirp, 69
LO, 186, 202
phase noise, 202
local oscillator, 202
phase noise, 202
long
tail response, 70
lowpass filter, 12, 15

magnetic
domain, 177
material, 177, 178
gyromagnetic effect, 181
moment, 177
magnetron, 216
Matthaei, 27
maximally flat, 18
MDS, 152
MEMS, 172, 175
switch, 176
microstrip
circulator, 179
filter, 79
isolator, 179, 180
microwave
modules, 1, 165
vacuum
device, 216
tube, 216
minimum
detectable signal, 152
discernible signal, 152
mixer, 186
analysis, 187
balanced, 191–193, 197
bilateral, 191
commutating, 195, 197
conversion loss, 189
diode, 187, 192
ring, 191–193
single ended, 191
double balanced, 197
triple balanced, 197
Gilbert, 198
image, 190
rejection ratio, 190
image reject, 197
noise, 190
figure, 189
performance parameters, 189
RFIC, 198
subharmonic, 196
switching, 194
terms, 189
triple balanced, 197
unbalanced, 197
waveforms, 191
module, 1, 165
design
budget method, 231
contribution method, 232
multipactor effect, 157
multiplier, 206
frequency, 207
parametric, 208
reactive, 208
resistive, 207

neodymium, 177
network identity
Kuroda's, 54
Norton's, 54, 56
NF, 131
mixer, 189
noise, 123, 128, 129
$1/f$, 202
temperature
effective, 128
cascade, 233
cascaded stages, 134
cascaded system, 133
correlated, 128
correlation coefficient, 128
environmental, 128
excess, 130, 131
factor, 131, 134, 189
system, 133
figure, 131
amplifier model, 139
measurement, 135
measurement, low noise, 137

measurement, Y-factor method, 135
mixer, 189
system, 133
two-ports, 139
flicker, 202
physical origin, 130
folding, 190
Friis's formula, 134, 233
measures, 130
mixer, 190
observation
frequency domain, 141
output of a two-port, 132
phase, 202
VCO, 205
PSD, 128
room temperature resistor, 126
shot, 129
physical origin, 129
single sideband, 204
source, 136
temperature, 128
measurement, 137
thermal
capacitor, 129
physical origin, 126
value at room temperature, 205
uncorrelated, 128
white, 128
Y-factor, 135
definition, 136
nonlinear
distortion, 144, 225
Norton's
identity, 56
transformation, 56
numerically controlled oscillator, 213

OIP3, 152, 229
cascaded system, 228, 234
operational amplifier, 67
organized cascade intercept method, 229
oscillator
carrier, 203
figure of merit, 204
local, 202
phase noise, 202
noise, 202
numerically controlled, 213
phase noise, 202
quadrature VCO, 205
QVCO, 205
VCO, 205
figure of merit, 204, 205
FOM, 204, 205
voltage-controlled, 204
OTA, 67
output
intercept point, 152
noise, 132

parallel
coupled-line
filter inhomogeneous medium, 114
parametric frequency multiplier, 208
passband ripple, 22, 31
passive
intermodulation distortion, 154
PBR, 22
PCL
filter in inhomogeneous medium, 114
PCL filter, 79
peak-to-mean envelope power ratio, 144
PFD
phase-frequency detector, 207
phase
accumulator, 213
comparator, 206
detector, 206
gain factor, 210
distortion, 144
frequency detector, 207
locked loop, 209
modulator, 212
noise, 202
oscillator, 202
VCO, 205
PIM, 154
PIN diode, 169, 175
switch, 175
PLL, 208, 209
amplitude demodulator, 212
applications, 211
capture range, 210
first-order, 210
frequency demodulator, 211
phase detector
gain factor, 210
phase modulator, 212
synthesizer, 211
VCO, 210
PMEPR, 144
back-off, 144
pole-zero description, 17
power
spectral density
noise, 128
prototype, 17
PSD
noise, 128
pump, 186

radar
duplexer, 180
radiometer, 138
rare earth elements, 177
R_{dB}, 22, 31
reactive frequency multiplier, 207, 208
rectifier, 167
remote sensing, 138
resistor
noise at room temperature, 126
resonator
LC-stub equivalence, 95
RF, 186
modules, 1, 165
RFIC
mixer, 198
Richards'
transformation, 53
highpass, 53
Richards' transformation, 50
lowpass, 52
ring mixer
transistor, 195, 200
ripple, 22, 31
factor, 22, 31
room temperature
noise of a resistor, 126

samarium, 177
satellite, 138
Scanlan, 27
schematic symbols, 5
Schottky
diode, 167
secondary electron emission, 157
SEY, 157
SFDR, 151–153
shot noise, 129
signal
-to-noise ratio, *see* SNR
two-tone, 146
singly terminated network, 13
SNR, 130, 152, 153
$\mathrm{SNR}_{\mathrm{MIN}}$, 152
sources, 214
SP4T switch, 175
SPDT switch, 175
spectrum
analyzer, 181
spin, 177
SPST switch, 175
spurious free dynamic range, 151
spurious passband, 114
standard
temperature, 131
stub
capacitively-loaded, 95
lumped resonator equivalence, 95
subharmonic mixer, 196
switch, 172, 175
diode, 175
FET, 175
MEMS, 172, 175, 176
pHEMT, 172, 175
PIN, 175
PIN diode, 172, 175
switching
mixer, 194
synthesis
ladder, 24, 26
synthesizer, 211
direct digital, 213

T_0, 131
tank circuit, 68
temperature
noise, 128
tensor, 178
third
-order intercept, 146
TPR, 14
transconductance amplifier, 67
transducer
function, 14
power ratio, 14
transferred electron device, 215
transit time device, 215
transmission
coefficient, 14
TRAPATT, 215
traveling wave tube, 216
triple balanced mixer, 197
tube
backward-wave oscillator, 216
travelling wave, 216
tuning
constant, 204
gain, 205
tunnel diode, 214
TUNNETT diode, 215
two-tone signal, 146
TWT, 216

UE, 33

unbalanced
  mixer, 197
unit
  element, 33
unorganized cascade
    intercept method, 230
up-conversion, 186

vacuum
  device, 216
    backward-wave
      oscillator, 216
    klystron, 216
    magnetron, 216
    travelling wave tube, 216
    TWT, 216
    TWTA, 216
  tube, 216
    TWT, 216
    TWTA, 216
varactor diode, 168
VCO, 204
  phase noise, 205
  PLL, 210
voltage
  controlled oscillator, 204
    in phase-locked loop, 210

white noise, 128
WiFi, 12

Y-factor
  definition, 136
  method, 135
YIG, 181
  filter, 179
  tuned filter, 181
yttrium-iron-garnet, 181

Zener diode, 169

MICROWAVE AND RF DESIGN: MODULES

Third Edition

Microwave and RF Design: Modules focuses on the design of systems based on microwave modules. The use of modules has become increasingly important in RF and microwave engineering for rapidly realizing high performance microwave systems. When integration is ultimately to be used, building a system up using modules provides a rapid means of prototyping and testing system concepts. A wide variety of RF modules including amplifiers, local oscillators, switches, circulators, isolators, phase detectors, frequency multipliers and dividers, phase-locked loops, and direct digital synthesizers are considered. Detailed design strategies for synthesizing filters based on parallel coupled lines are presented. The reader will gain an appreciation of design by synthesis. This book is suitable as both an undergraduate and graduate textbook, as well as a career-long reference book.

KEY FEATURES

- The fourth volume of a comprehensive series on microwave and RF design
- Open access ebook editions are hosted by NC State University Libraries at: https://repository.lib.ncsu.edu/handle/1840.20/36776
- 23 worked examples
- An average of 21 exercises per chapter
- Answers to selected exercises
- 6 case studies illustrating design procedures
- Emphasis on synthesis as well as building a rich library of microwave functions
- A companion book, *Fundamentals of Microwave and RF Design*, is suitable as a comprehensive undergraduate textbook on microwave engineering

ABOUT THE AUTHOR

Michael Steer is the Lampe Distinguished Professor of Electrical and Computer Engineering at North Carolina State University. He received his B.E. and Ph.D. degrees in Electrical Engineering from the University of Queensland. He is a Fellow of the IEEE and is a former editor-in-chief of *IEEE Transactions on Microwave Theory and Techniques*. He has authored more than 500 publications including twelve books. In 2009 he received a US Army Medal, "The Commander's Award for Public Service." He received the 2010 Microwave Prize and the 2011 Distinguished Educator Award, both from the IEEE Microwave Theory and Techniques Society.

OTHER VOLUMES

Microwave and RF Design
Radio Systems
Volume 1
ISBN 978-1-4696-5690-8

Microwave and RF Design
Transmission Lines
Volume 2
ISBN 978-1-4696-5692-2

Microwave and RF Design
Networks
Volume 3
ISBN 978-1-4696-5694-6

Microwave and RF Design
Amplfiers and Oscillators
Volume 5
ISBN 978-1-4696-5698-4

ALSO BY THE AUTHOR

Fundamentals of Microwave and RF Design
ISBN 978-1-4696-5688-5

Published by NC State University

Distributed by UNC Press

NC STATE UNIVERSITY